Astronomy & Astrophysics manuscript no. manuscript
June 11, 2018
©ESO 2018

Measuring precise radial velocities and cross-correlation function line-profile variations using a Skew Normal density⋆

U. Simola[1] ⋆⋆, X. Dumusque[2] ⋆⋆⋆, and Jessi Cisewski-Kehe[3]

[1] Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland
[2] Observatoire de Genève, Université de Genève, 51 ch. des Maillettes, CH-1290 Versoix, Switzerland
[3] Department of Statistics and Data Science, Yale University, New Haven, CT, USA

Received XXX; accepted XXX

ABSTRACT

Context. Stellar activity is one of the primary limitations to the detection of low-mass exoplanets using the radial-velocity (RV) technique. Stellar activity can be probed by measuring time dependent variations in the shape of the cross-correlation function (CCF), often estimated using different parameters of the modeled CCF. Therefore estimating the moments of the CCF with high precision is essential to de-correlate the signal of an exoplanet from spurious RV signals originating from stellar activity.
Aims. We propose to estimate the parameters of the CCF by fitting a Skew Normal (SN) density which, unlike the commonly employed Normal density, includes a skewness parameter to capture the asymmetry of the CCF induced by stellar activity and also the natural asymmetry induced by convective blueshift.
Methods. The performances of the proposed method are compared to the commonly employed Normal density, using both simulations and real observations, with different levels of activity and signal-to-noise ratio (SNR) levels.
Results. When considering real observations, the correlation between the RVs and the asymmetry of the CCF and the correlation between the RVs and the width of the CCF are stronger when using the parameters estimated with the SN rather than the ones obtained with the commonly employed Normal density. In particular the strongest correlations have been obtained when using as RV estimate the mean of the SN. This suggests that the asymmetry of the CCF and the width of the CCF estimated using a SN density may be more sensitive to stellar activity, which can be helpful when estimating stellar rotational periods and generally for characterizing stellar activity signals. The estimated uncertainties in the measured RVs using the proposed SN approach and as RV estimate the median of the SN are on average 10% smaller than the uncertainties calculated on the mean of the Normal. The estimated uncertainties on the asymmetry parameter of the SN are on average 15% smaller than the uncertainties measured on the Bisector Inverse Slope Span (BIS SPAN), which is the commonly parameter used to evaluate the asymmetry of the CCF.
Conclusions. We strongly encourage the use of the SN distribution to retrieve the different parameters of the CCF as the correlations used to probe stellar activity are stronger and the measure of the RV and the CCF asymmetry are more precise.

Key words. techniques: radial velocities – planetary systems – stars: activity – methods: data analysis

1. Introduction

When working with radial-velocities data (RVs), one of the main limitations to the detection of small-mass exoplanets is no longer the precision of the instruments used, but the different sources of variability induced by the stars (e.g. Feng et al. 2017; Dumusque et al. 2017; Rajpaul et al. 2015; Robertson et al. 2014). Stellar oscillations, granulation phenomena, and stellar activity can all induce apparent RV signals that are above the meter-per-second precision (e.g. Saar & Donahue 1997; Queloz et al. 2001; Desort et al. 2007; Dumusque et al. 2011; Dumusque 2016) reached by the best high-resolution spectrographs (HARPS, HARPS-N, Mayor et al. 2003; Cosentino et al. 2012). It is therefore mandatory to better understand stellar signals and to develop methods to correct for them, if in the near future we want to detect or confirm an Earth-twin planet using the RV technique. This is even more true now that instruments like the Echelle SPectrograph for Rocky Exoplanet and Stable Spectroscopic Observations (ESPRESSO) (Pepe et al. 2014) and the EXtreme PREcision Spectrometer (EXPRES) (Fischer et al. 2016) should reach the stability to detect such signals. However, if solutions are not found to mitigate the impact of stellar activity, the detection or confirmation of potential Earth-twins will be extremely challenging and false detections could plague the field.

One of the most challenging stellar signal to characterize and to correct for is the signal induced by stellar activity. Stellar activity is responsible for creating magnetic regions on the surface of stars, and those regions change locally the temperature and the convection, which can induce spurious RVs variations (e.g. Meunier et al. 2010; Dumusque et al. 2014; Borgniet et al. 2015). In theory, it should be easy to differentiate between the pure Doppler-shift induced by a planet, which shifts the entire stellar spectrum, and stellar activity, which modifies the shape of spectral lines and by doing so create a spurious shift of the stellar spectrum (Saar & Donahue 1997; Hatzes 2002; Kurster et al. 2003; Lindegren & Dravins 2003; Desort et al. 2007; Lagrange et al. 2010; Meunier et al. 2010; Dumusque et al. 2014). However, on quiet GKM dwarfs, the main target for precise RVs measurements, stellar activity can induce signals of a few $\mathrm{m\,s^{-1}}$. This corresponds physically to variations smaller than 1/100th of

⋆ Based on observations collected at the La Silla Parana Observatory, ESO (Chile), with the HARPS spectrograph at the 3.6-m telescope.
⋆⋆ e-mail: umberto.simola@helsinki.fi
⋆⋆⋆ Branco Weiss Fellow–Society in Science (url: http://www.society-in-science.org)

a pixel on the detector making the changing shape of the spectral lines challenging to detect. In order to measure such tiny variations, a common approach is to average the information of all the lines in the spectrum by cross correlating the stellar spectrum with a synthetic or an observed stellar template (Baranne et al. 1996; Pepe et al. 2002; Anglada-Escudé & Butler 2012). The result of this operation gives us the cross-correlation function (CCF). To measure the Doppler-shift between different spectra and therefore to retrieve the RVs of a star as a function of time, the variations of the CCF barycenter are calculated. The barycenter is generally estimated by fitting a Normal density to the CCF and retaining its mean. Variations in line shape between different spectra, which indicate the presence of signals induced by stellar activity, are measured by analyzing the different moments of the CCF. Usually, the width of the CCF is estimated using the full-width half-maximum (FWHM) of the fitted Normal density, and its asymmetry by calculating the CCF bisector and measuring the bisector inverse slope span (BIS SPAN, Queloz et al. 2001).

If an apparent RV signal is induced by activity, generally a strong correlation will be observed between the RV and chromospheric activity indicators like $\log(R'_{HK})$ or H-α (Boisse et al. 2009; Dumusque et al. 2012; Robertson et al. 2014), but also between the RV and the FWHM of the CCF or its BIS SPAN (Queloz et al. 2001; Boisse et al. 2009; Queloz et al. 2009; Dumusque 2016). It is therefore common now, that when fitting a Keplerian signal to a set of RVs to look for a planet, the model includes in addition linear dependancies with the $\log(R'_{HK})$, the FWHM and the BIS SPAN (Dumusque et al. 2017; Feng et al. 2017). It is also common to add a Gaussian process to the model to account for the correlated noise induced by stellar activity. The hyperparameters of the Gaussian process can be trained on different activity indicators (Haywood et al. 2014; Rajpaul et al. 2015) or directly on the RVs (Faria et al. 2016). It is therefore essential for mitigating stellar activity to obtain activity indicators that are the most correlated with the RVs but also for which we can obtain the best precision.

Several indicators have been developed that are more sensitive to line asymmetry than the BIS SPAN. In Boisse et al. (2011), the authors develop V_{span}, which is the difference between the RV measured respectively by fitting a Normal density to the upper and the bottom part of the CCF. This CCF asymmetry parameter is shown to be more sensitive than the BIS SPAN at low signal-to-noise ratio (SNR). Figueira et al. (2013) studied the use of new indicators, BIS-, BIS+, bi-Gauss and V_{asy}. The authors were able to show that when using bi-Gauss, the amplitude in asymmetry is 30% larger than when using BIS SPAN, therefore allowing the detection of lower levels of activity. They also demonstrated that V_{asy} seems to be a better indicator of line asymmetry at high SNR, as its correlation with RV is more significant than any of the previously proposed asymmetry indicators.

In all the methods described above, except bi-Gauss, the RV and the FWHM are derived using a Normal density fitted to the CCF, and the asymmetry is estimated using another approach. In this paper we propose to use a Skew Normal (SN) density to estimate with a single fit of the CCF, the RV, the FWHM and the asymmetry of the CCF, as this function includes a skewness parameter (Azzalini 1985).

The paper is organized as follow. In Sec. 2 we introduce the SN density, describe its applicability for modeling the CCF, and study how the SN parameters relate to the RV, FWHM and BIS SPAN of the CCF. In Sec. 3 we propose an expanded linear model to correct for stellar activity signals in RVs, which extends the linear models previously proposed for this purpose (e.g. Dumusque et al. 2017; Feng et al. 2017). In Sec. 4 the performance of the SN fit to the CCF is investigated using simulations coming from the Spot Oscillation And Planet 2.0 code (SOAP 2.0, Dumusque et al. 2014), followed by an analysis of real observations in Sec. 5. Sec. 6 considers derived error bars for the different estimated CCF parameters, and finally a discussion of the results and conclusions are included in Secs. 7 and 8, respectively.

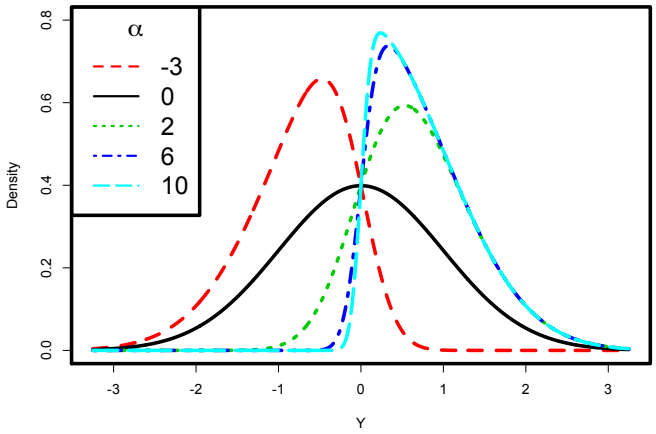

Fig. 1. Density function of a random variable Y following the SN distribution $SN(\xi, \omega^2, \alpha)$ with location parameter $\xi = 0$, scale parameter $\omega = 1$ and different values of the skewness parameter α indicated by different colors and line types. Note that the solid black line has an $\alpha = 0$ making it a Normal distribution.

2. The Skew Normal distribution

The Skew Normal (SN) distribution is a class of probability distributions which includes the Normal distribution as a special case (Azzalini 1985). The SN distribution has, in addition to a location and a scale parameter analogous to the Normal distribution's mean and standard deviation, a third parameter which describes the skewness (i.e. the asymmetry) of the distribution. Considering a random variable $Y \in \mathbb{R}$ (where $\mathbb{R}$ is the real line) which follows a SN distribution with location parameter $\xi \in \mathbb{R}$, scale parameter $\omega \in \mathbb{R}^+$ (i.e., the positive real line), and skewness parameter $\alpha \in \mathbb{R}$, its density at some value $y \in Y$ can be written as

$$SN(y; \xi, \omega, \alpha) = \frac{2}{\omega}\phi\left(\frac{y-\xi}{\omega}\right)\Phi\left(\frac{\alpha(y-\xi)}{\omega}\right), \quad (1)$$

where ϕ and Φ are respectively the density function and the distribution function of a standard Normal distribution[1]. The skewness parameter α quantifies the asymmetry of the SN. Examples of SN densities under different skewness parameter values and the same location and scale parameters ($\xi = 0$ and $\omega = 1$) are displayed in Fig. 1. A usual Normal distribution is the special case of the SN distribution when the skewness parameter α is equal to zero[2]. For reasons related to the interpretation of the parameters in Eq. 1 and computational issues with estimating α

[1] A standard Normal distribution is a Normal distribution with a mean of 0 and a standard deviation of 1.

[2] This can be seen from Eq. 1. If $\alpha = 0$ then $\Phi\left(\frac{\alpha(y-\xi)}{\omega}\right) = \Phi(0) = 0.5$ and therefore $SN(y; \xi, \omega, 0) = \frac{1}{\omega}\phi\left(\frac{y-\xi}{\omega}\right)$ which is the density of a Normal distribution. Note that $\Phi(0) = 0.5$ because $\Phi(0)$ is the the probability that a standard Normal random variable is less than or equal than 0.

near 0, a different parametrization is used in this work, which is referred to as the *centered parametrization* (CP). This CP is much closer to the parametrization of a Normal distribution, as it uses a mean parameter μ, a variance parameter σ^2 and a skewness parameter γ. In order to define the CP, we need to express the CP parameters (μ, σ^2, γ) as a function of (ξ, ω^2, α). This can be done using the following relations:

$$\mu = \xi + \omega\beta, \quad \sigma^2 = \omega^2(1-\beta^2), \quad \gamma = \frac{1}{2}(4-\pi)\beta^3\left(1-\beta^2\right)^{-3/2}, \quad (2)$$

where $\beta = \sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{\sqrt{1+\alpha^2}}\right)$ (e.g. Arellano & Azzalini 2010).

By using Eq. 2, the new set of parameters (μ, σ^2, γ) provides a clearer interpretation of the behavior of the SN distribution. For the α values used in Fig. 1, the corresponding values of (μ, σ^2, γ) are displayed in Table 1. In particular, μ and σ^2 are the actual mean and variance of the distribution, rather than simply a location and scale parameter, and γ provides an measure of the skewness of the SN. Along with the mean of the SN, we consider the median of the distribution as a measure of its barycenter. See Table 1 for the medians of the SN densities displayed in Fig. 1.

Table 1. CP values (μ, σ^2, γ) along with the median corresponding to the α values shown in Fig. 1, with location parameter $\xi = 0$ and scale parameter $\omega = 1$. Values are rounded to three decimal places.

α	μ	σ^2	γ	Median
-3	-0.757	0.427	-0.667	-0.672
0	0.000	1.000	0.000	0.000
2	0.714	0.491	0.454	0.655
6	0.787	0.381	0.891	0.674
10	0.794	0.370	0.956	0.674

Further details about the parametrization from Eq. 1, called the *Direct Parametrization* or DP, the CP, and general statistical properties of the SN are treated in rigorous mathematical and statistical viewpoints in the book by Azzalini & Capitanio (2014).

2.1. Fitting the Skew Normal density to the CCF

To fit the CCF using a SN density shape, we use a least-squares algorithm and the following model:

$$f_{CCF}(x_i) = \mathrm{C} - \mathrm{A} \times SN(x_i; \mu, \sigma^2, \gamma), \quad i = 1, \ldots, n \quad (3)$$

where C is an unknown offset for the continuum of the CCF, A is the unknown amplitude of the CCF, some times referred to as the CCF contrast, and μ, σ^2 and γ are the mean, variance and skewness of the SN as defined above. The values $x_1, \ldots, x_n$ are the different values of the x-axis of the CCF, generally in velocity units (e.g. $\mathrm{m\,s^{-1}}$).

When fitting a Normal density to the CCF, the estimated mean of the model is used as the estimated RV, the FWHM of the Normal density[3] represents the width of the CCF. Because the Normal density is symmetric, the skewness is always equal to 0 so a separate approach is needed to estimate the skewness of the CCF. An estimated skewness parameter is generally obtained by calculating the BIS SPAN of the CCF (see Sect. 1, and e.g. Queloz et al. 2001).

With the proposed SN approach, we propose two estimators of the RV: the mean and median of the SN model fit (referred to as SN mean RV and SN median RV, respectively), and present advantages and limitations for both of these choices in Sec. 5 and Sec. 6. The width of the SN, SN FWHM, is defined in the same way as for the Normal density[4], and finally the skewness of the CCF is estimated by the γ parameter.

To evaluate the strength of the correlation between the estimated RVs and the different stellar activity indicators, we calculated the Pearson correlation coefficient, R, which in its general form is defined as:

$$R(x, y) = \frac{cov(x, y)}{\sigma(x)\sigma(y)}, \quad (4)$$

where x and y are two quantitative variables, $cov(x, y)$ indicates the covariance between x and y, and $\sigma(x)$ and $\sigma(y)$ represent their standard deviations. A $p-$value for the statistical test having null hypothesis $H_0 : R = 0$ is provided, along with a 95% confidence interval for R when needed.

3. Radial Velocity correction for stellar activity

Exoplanets only produce a pure RV signal. On the contrary, stellar activity, in particular the presence of active regions on the stellar photosphere, do not produce blueshifts or redshifts of the entire stellar spectrum but can create spurious RV signals by modifying the shape of spectral lines. To track these variations in the shape of the spectral lines, the general approach consists in using the FWHM, the BIS SPAN or other indicators such as those introduced in Boisse et al. (2011) or Figueira et al. (2013), which provide information on the width and asymmetry of the CCF. A strong correlation between the estimated RVs and one or more of these parameters provides an indication that stellar activity signals may be affecting the measurements.

When fitting for planetary signals in RV data, it is common to include linear dependencies with the BIS SPAN and the FWHM to take into account the signal induced by stellar activity (e.g. Dumusque et al. 2017; Feng et al. 2017). We propose to add additional parameters in the model to correct for stellar activity: first the amplitude parameter A of the CCF, generally referred to as the CCF contrast, and the interaction between the BIS SPAN and the FWHM (or γ and SN FWHM in the SN case). The stellar activity correction we propose can therefore be written as:

$$RV_{\text{activity}} = \beta_0 + \beta_1 A + \beta_2 \gamma + \beta_3 \text{SN FWHM} + \beta_4(\gamma \text{SN FWHM}) + \epsilon, \quad (5)$$

where β_0 is the intercept and ϵ is the error with mean equal to 0 and covariance matrix equal to $\sigma^2 I$ (I defined as the identity matrix). The contrast parameter A accounts for the presence of a spot on the stellar surface, which produces a change in the amplitude of the CCF and not only on its asymmetry or width (see e.g. Fig. 2 in Dumusque et al. 2014). The benefits of including a variable that quantifies the interaction between γ and SN FWHM (or BIS SPAN and FWHM) will be better understood through the results of the examples presented in Sec.4. This interaction term can account for possible interactions between SN FWHM (or FWHM) and γ (or BIS SPAN), meaning that each variables' association with the response, RV_{activity}, depends also on the other variable.

The proposed model is analyzed using statistical tests on the parameters β_0, β_1, β_2, β_3 and β_4 where the null hypothesis is $H_0 : \beta_i = 0$, for $i = 0, \ldots, 4$. The significance level for the

[3] FWHM= $2\sqrt{2\ln 2}\,\sigma$ with standard deviation σ

[4] Note that SN FWHM does not correspond to the width of the SN density at half maximum like in the Normal case.

tests are set at 0.05. The coefficient of determination, R^2, is used to assess how well the proposed linear combination of variables accounts for the variability of RV_{activity}.

The proposed function defined in Eq. 5 is the result of statistical and astronomical considerations. In particular we checked that the correlations between the proposed parameters were not approaching one: if it was the case, the matrix needed to calculate the estimates would be singular, hence non invertible. This problem is known in statistics with the term multicollinearity. A detailed discussion of the topic can be found in the book by Belsley (1991). In the analysis of real data presented in this work, we never observed a correlation coefficient exceeding 0.66 between the asymmetry and width parameters and therefore, the problem of multicollinearity is avoided. Note that we investigated the statistical significant of the interaction term between A and the width, and A and the asymmetry of the CCF, however, those interaction were not relevant for accounting for stellar signal.

4. Simulation Study

In order to evaluate the performance of the proposed SN approach for modelling the CCF and the benefit of using the proposed correction for stellar activity (See Eq. 5), we begin by considering a simulation study using spectra generated from the Spot Oscillation And Planet 2.0 code (SOAP 2.0, Dumusque et al. 2014).

For a given configuration of spots and faculae on the stellar surface, SOAP 2.0 gives as output the simulated CCF as a function of rotational phase. The code also returns the RV and the FWHM by fitting a Normal density to the CCF, and the BIS SPAN by calculating the bisector of the CCF. SOAP 2.0 gives noiseless CCFs affected by stellar activity, which are used to compare the benefits of a SN density fit to the CCF compared to a Normal density fit.

For the simulations discussed below, a star similar to the Sun was modeled, with a solar disc of one solar radius seen equator-on, and with a stellar rotational period set to 25.0 days. The stellar effective temperature is set to 5778 K , and a quadratic limb-darkening relation with linear and quadratic coefficients 0.29 and 0.34 are used, respectively (Oshagh et al. 2013; Claret & Bloemen 2011). In order to make the result of the simulations more comparable to real data obtained with the HARPS spectrograph discussed in Sect. 5, the SOAP 2.0 CCFs were generate with a width of 40 km s^{-1} and considering initial spectra with a spectral resolution of R=115'000.

4.1. Faculae

To see the impact of a facula on the different parameters of the CCF, we simulated the effect of an equatorial faculae of size 3% relative to the visible stellar hemisphere. The faculae is face-on when the phase equals to 0. Note that a 3% faculae is relatively large for the Sun; at maximum activity, big faculae have generally a size of 1% (e.g. Borgniet et al. 2015). In Fig. 2, we compare the barycentric variation of the CCF as measured when fitting a Normal density and using its mean (N mean RV), and when fitting a SN density and taking its mean (SN mean RV) or its median (SN median RV). We see that all the different estimates of the CCF barycenter present a signal of similar amplitude, however the signal obtained with SN mean RV is different from the two others with a maximum amplitude happening at a different phase.

Correlations between the different RV estimates and the different CCF asymmetry or width estimates are displayed in Fig. 3.

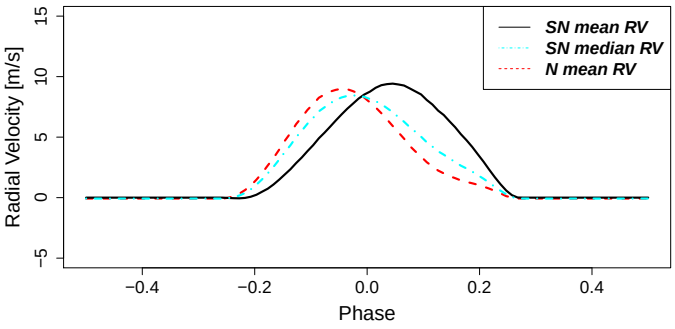

Fig. 2. RV estimates for N mean RV (red dashed line), SN mean RV (black line), and SN median RV (cyan dots-dashed line). In this case, the CCFs were generated using SOAP 2.0 with an equatorial 3% facula on the simulated Sun. The star does one full rotation between phase -0.5 and 0.5, with the facula being seen face-on for phase 0. The variations observed in SN mean RV are notably different from the variations measured in SN median RV and N mean RV.

The strength of the correlation between γ and SN mean RV, and γ and SN median RV are stronger than the correlations between BIS SPAN and RV, with Pearson correlation coefficient values of R=0.46, -0.67 and -0.09, respectively. For the width barycenter correlations, there is a stronger correlation between SN FWHM and SN mean RV compared to the one between FWHM and N mean RV, $R = 0.98$ and 0.84, respectively. In this case however, the correlation between SN FWHM and SN median RV is smaller with $R = 0.50$. This first analysis shows that in the case of a facula, using some parameters from the SN can lead to stronger correlation than the usual Normal parameters and therefore, the SN parameters may better probe stellar activity. We investigate this feature further in the next sections where we consider simulated data with a single spot and a spot plus a planet, and in Sec 5 with real observations.

Since the RV variation displayed in Fig. 2 is caused by only stellar activity, in this case a facula, we applied the activity correction proposed in Eq. 5 to check its performance in this setting. The results of this correction are displayed in Fig. 4 and the statistical tests on the coefficients involved in Eq. 5 are summarized in Table 2. The proposed correction for stellar activity is able to account for the majority of the activity signal created by a facula, with a R^2 of our model larger than 0.95. In addition, the rms of the different estimates of the RV reduces from about 3 m s^{-1}before correction to values below 0.15 m s^{-1}after correction. We see a slightly smaller rms after correction when using the SN parameters compared to the Normal parameters, however the different is probably not significant. When comparing the correction proposed in Eq. 5 with what is generally used (i.e. a linear combination of only the asymmetry and width parameter), we see that the proposed correction is able to reduce the rms of the RV residuals by a factor of 2. Looking at the significance of the coefficients in table 2, we observe that all the Sn or Normal parameters are relevant for the correction.

4.2. Spot

In this section, we consider the effects on the CCF parameters of an equatorial spot of size 1% relative to the visible stellar hemisphere. The spot is face on when the phase equals to 0. Note that this is a large spot for the Sun, as in general large spots are more in the regime of 0.1% (e.g. Borgniet et al. 2015). In

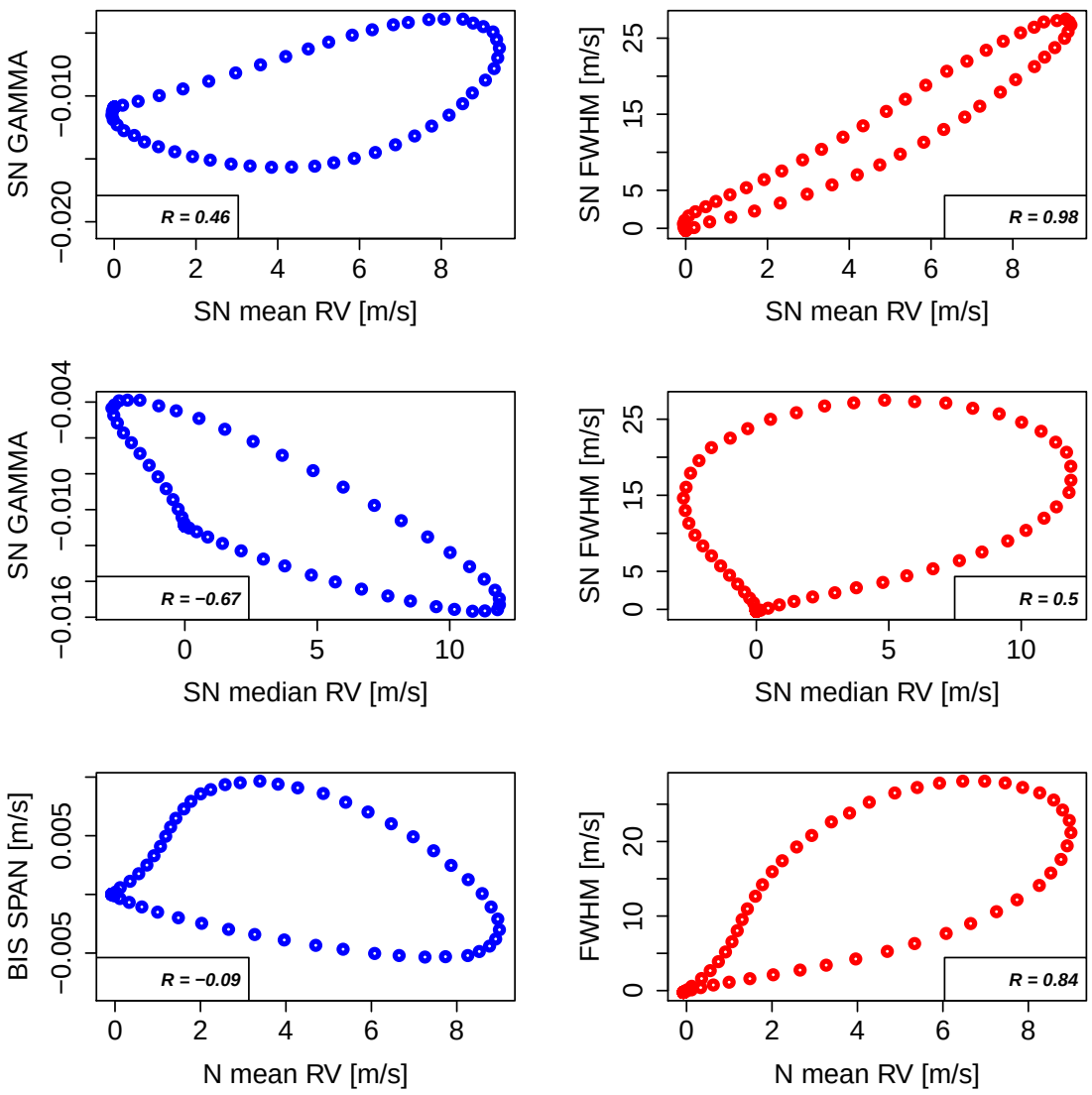

Fig. 3. (left) Correlations between the different asymmetry parameters and their corresponding RV estimates in the case of an equatorial 3% facula on the simulated Sun. (right) Correlations between the different width parameters and their corresponding RV estimates for the same facula. In the presence of a facula, both the shape and the width of the CCF change as the star rotates, producing statistically significant correlations.

Table 2. P–values for the estimated coefficients from the model in Eq. 5 for correcting stellar activity induced by an equatorial 3% facula on the simulated Sun. All the parameters corresponding to the Normal or SN variables are statistically significant to explain the spurious RV variations caused by the facula. The estimated R^2 show that the proposed correction for stellar activity explains the vast majority of the spurious variability present in the different RV estimates.

Parameter	N mean RV	SN mean RV	SN median RV
β_0	0.033	0.00020	0.61
β_1	$2.22e-16$	$2.22e-16$	$2.22e-16$
β_2	0.0034	$2.22e-16$	$2.22e-16$
β_3	0.00016	$1.091e-6$	$9.75e-7$
β_4	$2.22e-16$	$2.22e-16$	$2.22e-16$
R^2	0.9978	0.9985	0.9981

Fig. 5, we shows the barycentric variation of the CCF induced by this simulated spot. In contrast to the case of the facula, all the different estimates of the CCF barycenter for the spot have the same shape in variation. The amplitude for SN mean RV is however slightly smaller.

Fig. 6 shows the correlations between the asymmetry parameters and the different estimates for the CCF barycenter (i.e. SN mean RV, SN median RV and N mean RV). The correlation between γ and SN median RV is the strongest with a $R = 0.94$, followed by the correlation BIS SPAN - N mean RV and γ - SN mean RV, with $R = 0.86$. Regarding the correlation between the width and the CCF barycenter, we note that the variation is seen as a circle in this parameter space and therefore no correlation is observed. Once again, like in the case of the facula, we see that some parameters of the SN gives stronger correlations than the Normal parameters.

As before, the originally RVs were corrected by using Eq. 5. The results of the correction are displayed in Fig. 7 and in Table 3. Like for the faculae, the proposed correction is able, in

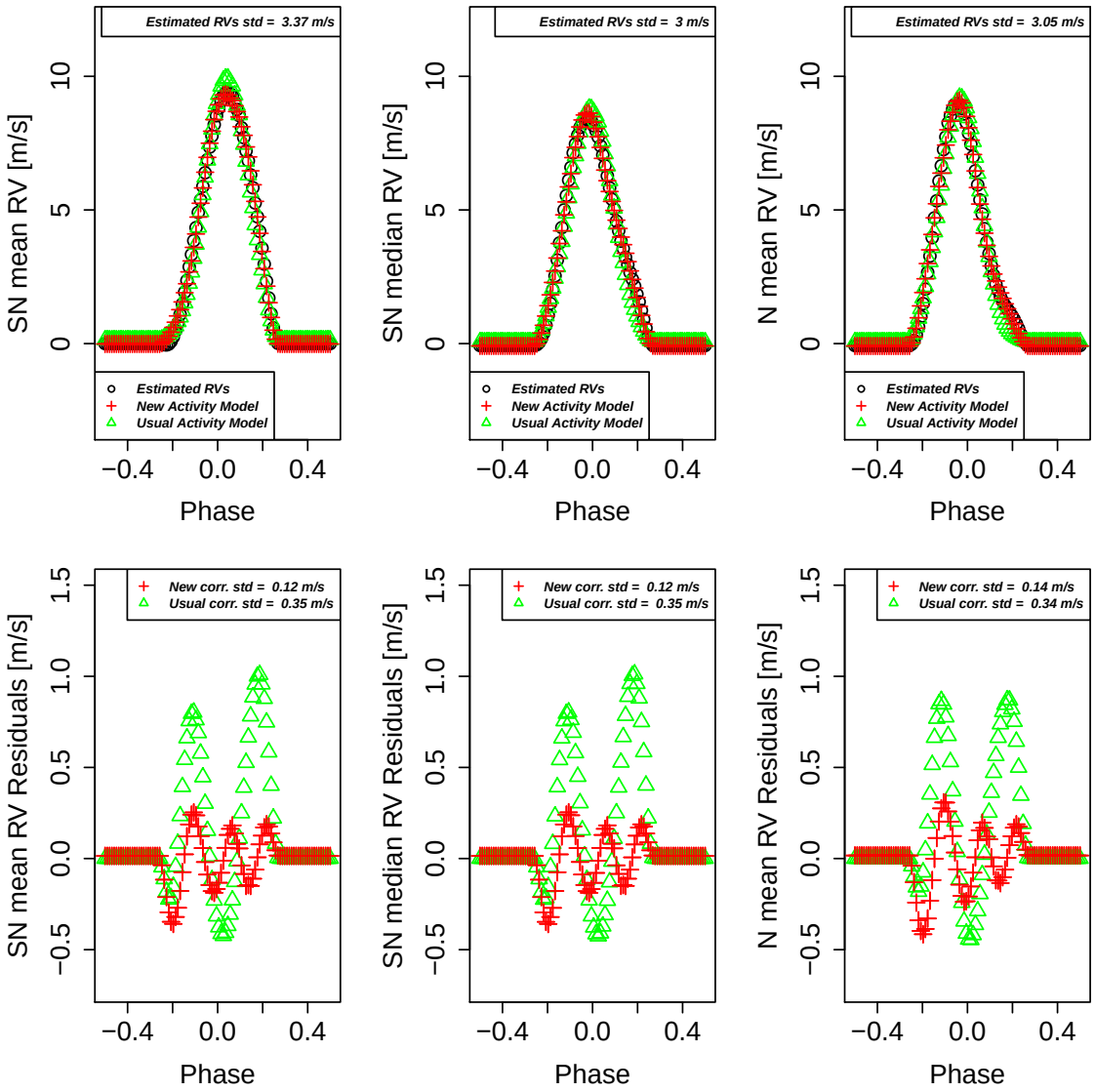

Fig. 4. (top) The spurious estimated RVs (black dots) caused by a facula in the simulated data using a Normal and a SN fit, the estimated RVs using Eq. 5 (red crosses), and the estimated RVs using the usual correction for stellar activity (green triangles), based on $RV_{\rm activity} = \beta_0 + \beta_1\gamma + \beta_2$SN FWHM for the SN fit and on $RV_{\rm activity} = \beta_0 + \beta_1$BIS SPAN $+ \beta_2$FWHM for the normal fit. (bottom) The residuals from the model fit using Eq. 5 (New corr. std, red crosses) and the residuals from the usual correction (Usual corr. std, green triangles). The residuals have a smaller systematic component when using the proposed model of Eq. 5 (red crosses) compared to the usual model (green triangles). The tests of statistical significance on the parameters are presented in Table 2.

the case of the spot, to almost model entirely the stellar activity signal when considering the SN or Normal parameters, with R^2 values for the linear combination larger than 0.99. Looking at Fig. 7, we see that the proposed activity correction is able to reduce the signal of a spot from a raw RV rms larger than 4.80 m s^{-1}down to a rms of 0.38 m s^{-1}, for any of the different RV estimates. When comparing the activity correction proposed in this paper with what is commonly used, i.e only a linear dependance with the width and asymmetry of the CCF, we see that our solution is capable of reducing the RV residual rms by a factor of 3.5, which is even more than the factor 2 found in the case of the faculae.

Looking at Table 3, we see that the Normal or SN parameters appearing in Eq. 5 are all statistically significant to explain the activity signal, except the width of the CCF when the RVs are derived with the mean or median of the SN density. This is not surprising when looking at the circle shape drawn when plotting the width as a function of the RV in Fig. 6.

4.3. Spot and planet

The final simulation presented here includes a planetary signal influencing the CCF along with the 1% spot modeled previously (see Sec. 4.2). The purpose of this example is to check if we are able to disentangle these two different sources of variations when using the parameters derived using a Normal versus a SN fit to the CCF. In this scenario the planet is injected with a semi-amplitude of 10 m s^{-1}with no eccentricity and with a period corresponding to one third of the stellar rotational period, i.e. one-third of 25 days.

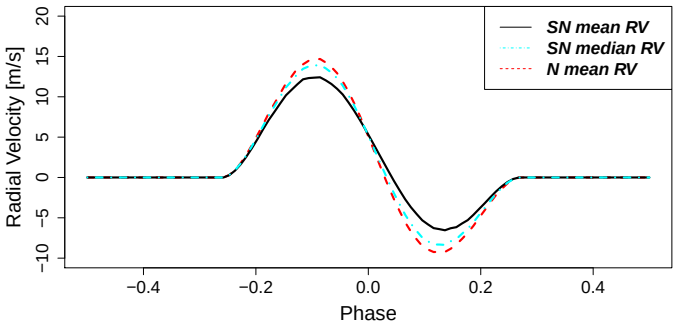

Fig. 5. RV estimates for N mean RV (red dashed line), SN mean RV (black line) or SN median RV (cyan dots-dashed line) using CCFs generated from SOAP 2.0 with an equatorial 1% spot on the simulated Sun. The star does one full rotation between phase -0.5 and 0.5, with the spot being seen face-on for phase 0. SN mean RV seems to have the smallest spurious variations caused by the spot.

Table 3. P–values for the different coefficients used in Eq. 5 for the correction of stellar activity induced by an equatorial 1% spot on the simulated Sun. All the parameters corresponding to the Normal or SN parameters are statistically significant to explain the spurious RV variations caused by this spot, except for the width of the CCF when using SN mean RV or SN median RV as RV estimates. The estimated R^2 show that the proposed correction for stellar activity explains the vast majority of the spurious variability seen in the different RV estimates.

Parameter	N mean RV	SN mean RV	SN median RV
β_0	0.4975	0.21	0.21
β_1	$2e-16$	$2e-16$	$2e-16$
β_2	$2e-16$	$2e-16$	$2e-16$
β_3	0.017	0.13	0.11
β_4	$2e-16$	$2e-16$	$2e-16$
R^2	0.9959	0.9936	0.9952

Fig. 8 shows the variation observed in the CCF barycenter parameters. As in the case of the spot, all RV estimates show similar variations, with SN mean RV showing a slightly smaller amplitude.

The correlation between the different CCF parameters are displayed in Fig. 9. The correlations are weaker than in the case of the spot due to the planet inducing changes in RV without affecting the width or asymmetry of the CCF. However, the strength of the correlations between the CCF asymmetry and RV are in the same order as with the spot-only model: γ–SN median RV has the highest correlation followed by BIS SPAN–N mean RV and then γ–SN mean RV, with R values of -0.84, -0.78 and -0.76, respectively. The patterns seen in the width-RV phase space in Fig. 9 follow a circle similar to the spot-only model, and no correlation is observed between those two parameters.

In order to correct the estimated RVs from the spurious variation caused by the spot, the proposed model for correcting the activity is added to a signal that takes into account the RV variation caused by a injected planet. The observed RV can therefore be modeled as a combination of the activity and planetary signals:

$$RV = RV_{\text{activity}} + RV_{\text{planet}}, \tag{6}$$

where RV_{activity} can be found in Eq. 5, and RV_{planet}, in the case with no eccentricity, can be modeled by the following sinusoidal function:

$$RV_{\text{exoplanet}} = K \sin\left(\frac{2\pi}{P}(t - t_0)\right), \tag{7}$$

with amplitude K, orbital period P, and an epoch at the periapsis t_0. The previous three unknown parameters define the planetary orbit.

Table 4. P–values for the different coefficients used in Eq. 5 for the correction of stellar activity induced by an equatorial 1% spot on the simulated Sun, and a planet with period one third of the rotational period and a semi-amplitude 10 m s^{-1}. All the parameters corresponding to the Normal or SN variables are statistically significant to explain the spurious RV variations caused by this spot plus planet, except for the width of the CCF. Note that since nonlinear least squares was required, the residual standard error rather than the R^2 is displayed as a reference. [[**Jessi:** is there a reason the actual model, (6), with the Doppler shift was not references?]]

Parameter	N mean RV	SN mean RV	SN median RV
β_0	0.00063	$2e-16$	$1.42e-09$
β_1	$2e-16$	$2e-16$	$2e-16$
β_2	$2e-16$	$2e-16$	$2e-16$
β_3	0.067	0.40	0.38
β_4	$2e-16$	$2e-16$	$2e-16$
K	$2e-16$	$2e-16$	$2e-16$
P	$2e-16$	$2e-16$	$2e-16$
t_0	$2e-16$	$2e-16$	$2e-16$
Residuals	0.71 m s^{-1}	0.66 m s^{-1}	0.70 m s^{-1}

The proposed model from Eq. 6 was fitted to the RV data and the results of the estimated model are summarized in Table 4. Except for the width parameters with coefficient β_3, all the other Normal or SN parameters are significantly useful to explain the RV variation induced by a spot plus a planet. We also observe that the RV residuals, once corrected for stellar activity and the presence of the planet, are comparable in terms of rms for all the three different RV estimates, with SN mean RV giving a slightly smaller value.

5. Real data application

In this Section we present the analysis conducted on the star Alpha Centauri B (Alpha Cen B), comparing the result of fitting a CCF using the SN density defined in Sec. 2.1 with the usual approach based on fitting a Normal density for estimating the RV and asymmetry of the CCF. Four other stars have been analyzed with the proposed method and details can be found in Appendix A. For all the stars considered in the presented work, only CCFs that were derived from spectra that had at least a SNR of 10 at 550 nm were selected.

5.1. Comparison for Alpha Cen B of the different CCF parameters derived with the Normal and the Skew Normal

There were 1808 CCFs analyzed that were derived from the spectra of Alpha Centauri B taken in 2010 by the HARPS spectrograph. Note that more observations were carried out this year, however only the data that were not significantly affected by contamination from Alpha Centauri A were used (see Dumusque et al. 2012). The selected observations represent probably, among all RV data existing, the best sampled and most

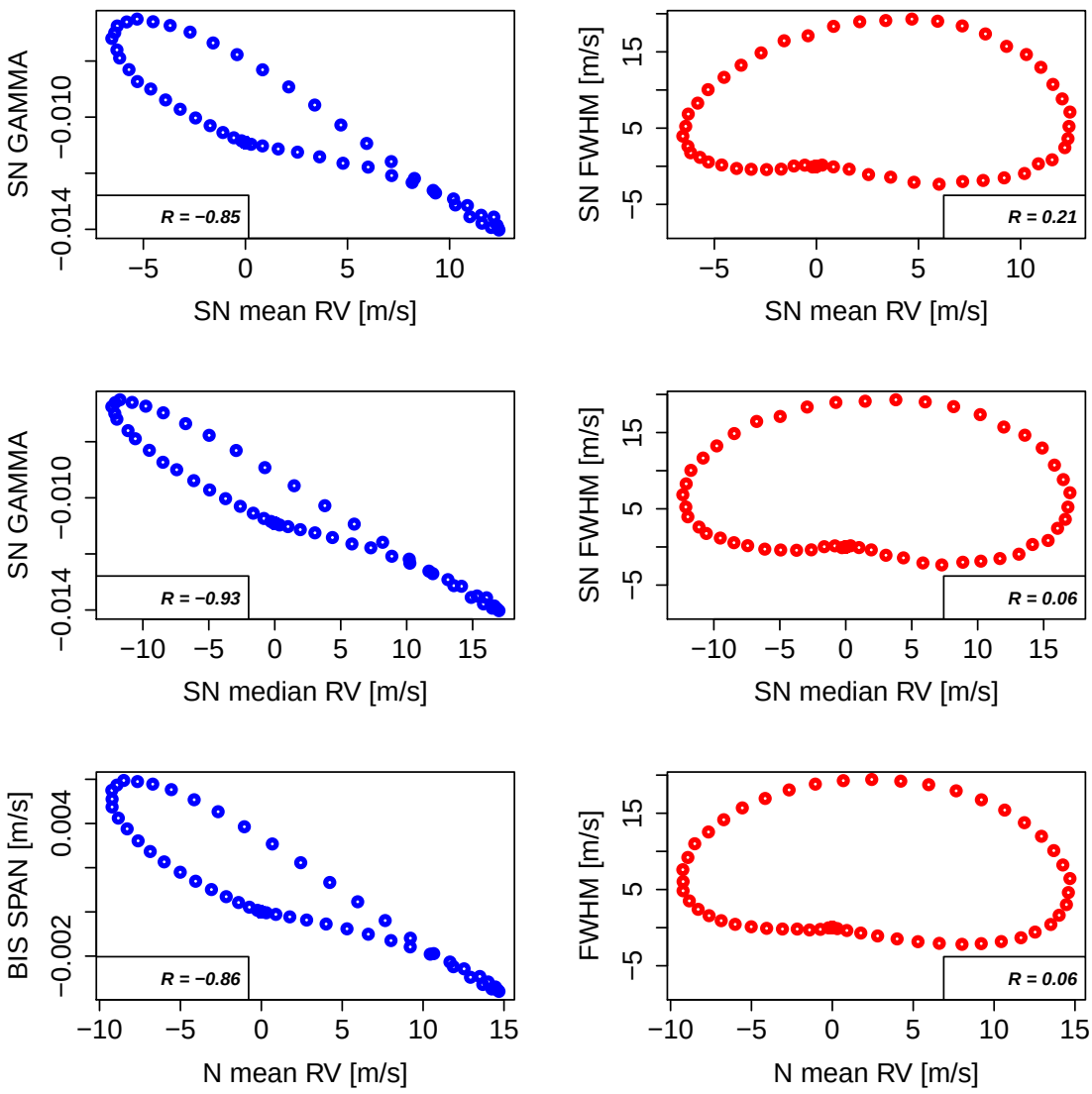

Fig. 6. (left) Correlations between the different asymmetry parameters and their corresponding RV estimates in the case of an equatorial 1% spot on the simulated Sun. (right) Correlations between the different width parameters and their corresponding RV estimates for the same spot. In the presence of a spot, both the shape and the width of the CCF change as the star rotates. However, only the asymmetry produces a statistically significant correlation with the different RV estimates. The width parameters and their corresponding RV estimates present weak correlations and in general much weaker correlations respect to the results obtained when an equatorial 3% facula is present on the simulated Sun.

precise RV data set showing strong solar-like activity signal (Thompson et al. 2017; Dumusque et al. 2012).

First, the correlation between γ and BIS SPAN is evaluated. In the left panel of Fig. 10, we see that the relationship between γ and the BIS SPAN is linear, with a slope equal to 720 and a strong Pearson correlation coefficient of $R = 0.95$. This strong correlation suggests that the γ and BIS SPAN are measuring a similar asymmetry for the CCF. This strong correlation allows as well to convert the dimensionless γ parameters into $\mathrm{m\,s^{-1}}$ using the slope of the correlation, in this case 720 $\mathrm{m\,s^{-1}}$.

The right plot of Fig. 10 displays the comparison between the RVs derived estimated using the SN density and the Normal density. The amplitude of the activity signal is slightly stronger for the SN mean RV (in the top-right plot the black circles of the SN mean RV tend to be more extreme), while the signal measured using N mean RV or SN median RV tend to be similar. This behavior is similar to the faculae simulated with SOAP 2.0 in see Sec. 4.1, suggesting that the activity signal could be due to faculae present on Alpha Centauri B. Another argument that suggests the presence of faculae, is the strong positive measured correlation between γ and SN mean RV and SN FWHM and SN mean RV, as displayed in Fig. 11. In Sec. 4.2, we saw that a spot induces a negative correlation between γ and SN mean RV, while weak correlations are measured between SN FWHM and SN mean RV. Dumusque (2014) had previously suggested that the activity of Alpha Centauri B could be due to faculae.

Similar to the analyses in Sec. 4, we compare the correlation between the asymmetry or the width parameters of the CCF and the RV in Fig. 11. For this analysis, we also include the asymmetry parameters derived in Boisse et al. (2011), V_{span} and in Figueira et al. (2013), BIS-, BIS+, Bi Gauss and V_{asy}, as these authors found those asymmetry parameters more correlated to the RVs than BIS SPAN. It is clear in the case of Alpha Cen B, that the correlation found between γ and SN mean RV is the

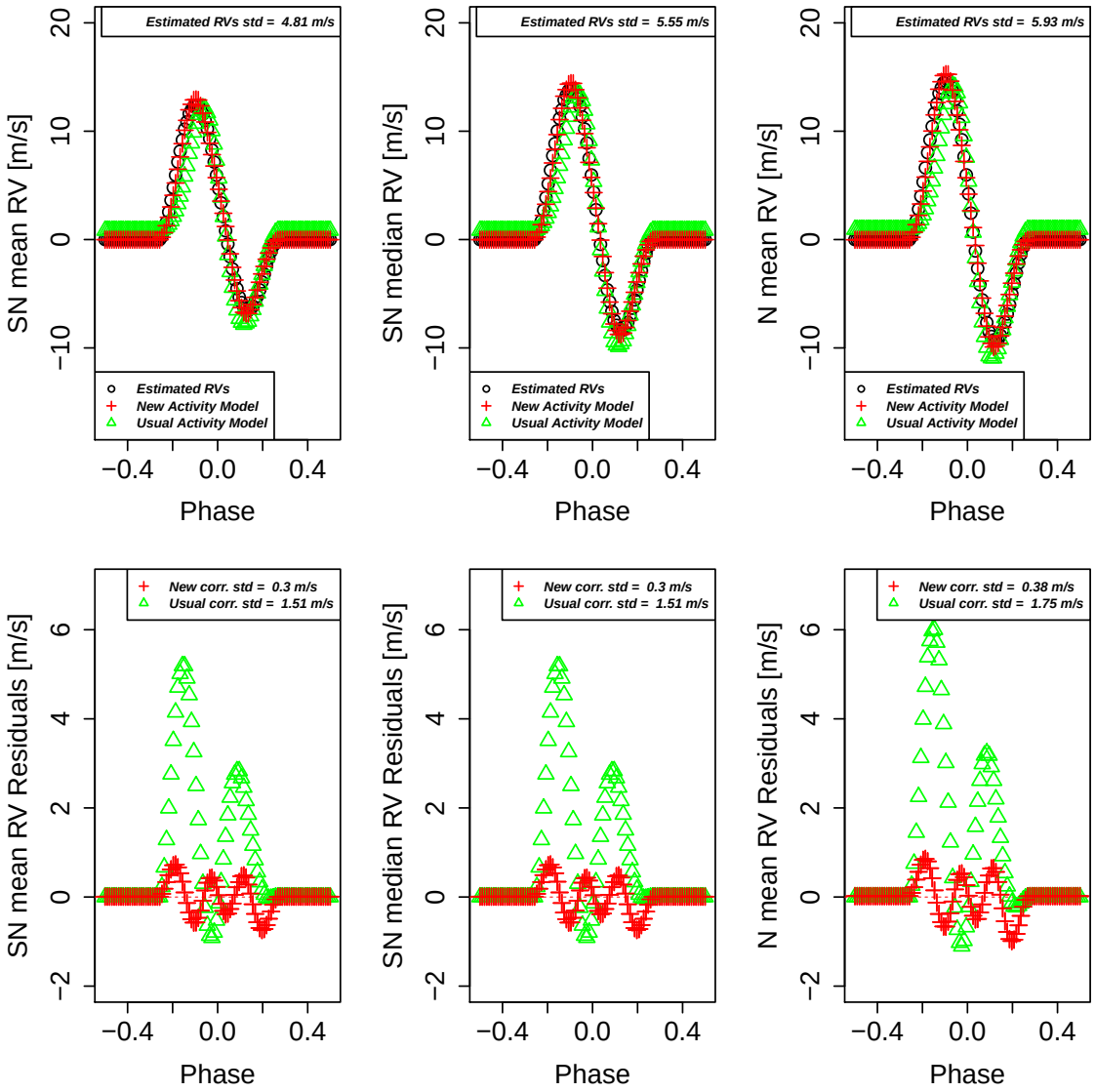

Fig. 7. (top) The spurious estimated RVs (black dots) caused by a spot in the simulated data, the estimated RVs using Eq. 5 (red crosses) and the estimated RVs using the usual correction for stellar activity (green triangles), based on $RV_{\text{activity}} = \beta_0 + \beta_1\gamma + \beta_2$SN FWHM for the SN fit and on $RV_{\text{activity}} = \beta_0 + \beta_1$BIS SPAN $+ \beta_2$FWHM for the normal fit. (bottom) The residuals from the model fit using Eq. 5 (New corr. std, red crosses) and the residuals from the usual correction (Usual corr. std, green triangles). The residuals have a smaller systematic component when using the proposed model of Eq. 5 (red crosses) compared to the usual model (green triangles). The tests of statistical significance on the parameters are presented in Table 3.

strongest. The Pearson correlation coefficient is $R = 0.74$, while it is at best $R = 0.42$ for all the other asymmetry–RV correlations not derived using the SN density fit. When looking at the correlations between the width and the RV estimates for Alpha Cen B, again the correlation is strongest for the SN parameters with $R = 0.82$ for SN FWHM-SN mean RV, while it is $R = 0.70$ for FWHM and N mean RV.

Results illustrating the performance of the stellar activity correction proposed in Sec. 3 is displayed in Fig 12. For Alpha Cen B, the RV estimated with SN mean RV has a std that is 35% larger than the std of the RV estimated with the N mean RV, and the std of SN median RV is 9% larger than that of the N mean RV. Even though we see these differences in the estimated RV, once we correct for the stellar activity using Eq. 5, the rms of the residuals are essentially the same for all three approaches. Although when studying the correlations between the different parameters we see that the parameters retrieved by using the SN density are more sensitive to stellar activity than the ones obtained with a Normal density fit, the proposed linear model for correcting from stellar activity is not able to perform better in the SN case than in the Normal case. When comparing the new correction for stellar activity proposed in Sec. 3 with the usual correction, that uses only a linear combination of the width and the asymmetry of the CCF, we end up with very similar results. The new correction is however able to get RV rms that are 6 cm s^{-1} smaller than the usual correction.

When looking at the statistical significance of the different parameters used for correcting activity in Table 5, we see that the intercept is not significant for any of the models, and BIS SPAN (coefficient β_2) does not statistically significantly help in explaining stellar activity when using the parameters derived from the Normal density fit. However all the other parameters

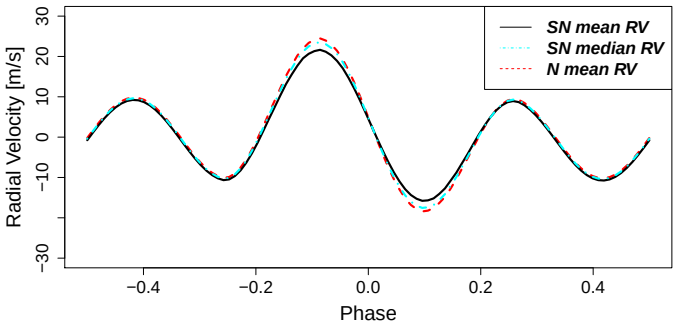

Fig. 8. RV estimates for N mean RV (red dashed line), SN mean RV (black line) or SN median RV (cyan dots-dashed line). In this case, the CCFs have been generated using SOAP 2.0, considering an equatorial 1% spot on the simulated Sun in addition to a planet with a period of one third of the rotational period of the star and with an amplitude of 10 m s^{-1}. The star do one full rotation between phase -0.5 and 0.5, with the spot being seen face-on for phase 0.

Table 5. P–values for the different coefficients used in Eq. 5 for the correction from stellar activity in Alpha Cen B data. All the variables corresponding to the Normal or SN parameters are statistically significant, except for the asymmetry of the CCF when using the BIS SPAN with the N mean RV. The evaluation of the R^2 shows that the proposed linear combination better explains variations in RVs due to the stellar activity when the RVs are derived using SN mean RV.

Parameter	N mean RV	SN mean RV	SN median RV
β_0	0.49	0.90	0.027
β_1	$2.22e-16$	$2.22e-16$	$2.22e-16$
β_2	0.33	$2.22e-16$	$1.23e-11$
β_3	$2.22e-16$	$2.22e-16$	$2.22e-16$
β_4	$2.22e-16$	$2.22e-16$	$2.22e-16$
R^2	0.57	0.78	0.66

in the Normal and SN cases are statistically significant for modeling stellar activity. By analyzing the values of the coefficient of determination, R^2, we see that the model for the SN mean RV is able to capture the highest percentage of variability in the estimated RV. This is not a surprising result since for the three different RV estimates, we arrive to the same RV residual rms after correction for activity, but before correction, SN mean RV shows the largest RV rms (see Fig 12).

5.2. Comparison for HD192310, HD10700, HD215152 and Corot-7 of the different CCF parameters derived with the Normal and the Skew Normal

In the previous section we evaluated the improvement obtained by the SN parameters compared to the Normal parameters and the BIS SPAN. We carryout similar analyses for four other main sequence stars: HD192310, HD10700, HD215152 and Corot-7. The same correlations plots and activity–after–correction plots displayed in the previous section for Alpha Cen B can be found found for those new four stars in Appendix A.

The correlations between the parameters of these additional stars are similar to those obtained for Alpha Cen B. The correlation between γ and SN mean RV is the strongest among all the asymmetry-RV correlations. For the correlation between the width and the estimated RV, the strongest correlation is consistently between SN FWHM and SN mean RV. However, there is one exception in the case of HD10700 where the Pearson correlation coefficient between FWHM and N mean RV is equal to $R = 0.53$, while it is $R = 0.42$ between SN FWHM and SN mean RV.

Except for the special case discussed above for HD10700, the analyses of those four stars, in addition to the analyses on Alpha Cen B, shows that the parameters derived when using a SN density are more sensitive to activity. Therefore using the SN parameters, and in particular defining RV as SN mean RV, can lead to better detection of stellar activity over the Normal parameters. This is the case for the evaluation of the asymmetry-RV correlations for Alpha Cen B, HD10700, HD215152, HD192310 and Corot-7, and the width-RV correlation for Alpha Cen B, HD215152, HD192310 and Corot-7 (see Appendix A).

In terms of correction for stellar activity, in the case of Alpha Cen B we see that although the RV rms is larger when measuring the RVs using SN mean RV (compared to the RVs obtained using N mean RV), once corrected for activity using the new model proposed in Sec. 3, both RVs estimates provides similar residuals. We also observe slight differences in the rms between the new proposed correction and the usual correction, which uses only a linear combination of the width and the asymmetry of the CCF. For Corot-7 the new correction is able to provide RV residual rms 30 cm s^{-1}smaller than the one obtained with the usual correction.

5.3. Detection limits when using the estimated RVs from the Normal or the Skew Normal models

In the previous section, we saw that the estimated RV measured when considering a SN or a Normal model resulted in different amplitudes, especially when using the SN mean RV. However, once corrected for stellar activity using the linear combination presented in Eq. 5, as shown in the bottom plots of Fig 12, the rms of the residuals are essentially the same for all three approaches. In this section, we investigate the ability of the three different RV estimators' (N mean RV, SN mean RV, and SN median RV) ability to detect planetary signals among stellar activity. To carryout this test, the minimum detected amplitude of a planetary signal is estimated at different orbital periods when considering data affected by stellar activity.

In order to obtain CCFs affected by realistic stellar activity signals, the CCFs from Alpha Cen B used previously were considered. To simulate a planetary signal, the CCFs were blue- or red-shifted with the desired amplitude, period, and phase. Several RV data sets with the same stellar signal, but different planetary signals were injected using parameters corresponding to the following grid:

- period of 3, 5, 7, 9, 11, 15, 20, 25, and 30 days,
- amplitude from 0.5 to 3 m s^{-1} by steps of 0.05 m s^{-1},
- 10 different phases, evenly sampled between 0 and 2π.

For each of the 4500 simulations considered, the activity correction presented in Eq. 5 was applied, and signals in the residuals were investigated using a Generalized Lomb-Scargle periodogram (Lomb 1976; Scargle 1982; Zechmeister & Kürster 2009). If a signal with a p–value[5] smaller than 1% had a period compatible with the injected planetary period within an error budget of 20%, the signal was considered significant and the corresponding planet considered detected. Finally, for each period considered, the minimum amplitude at which at least 50%

[5] The p–values were estimated using a bootstrap procedure.

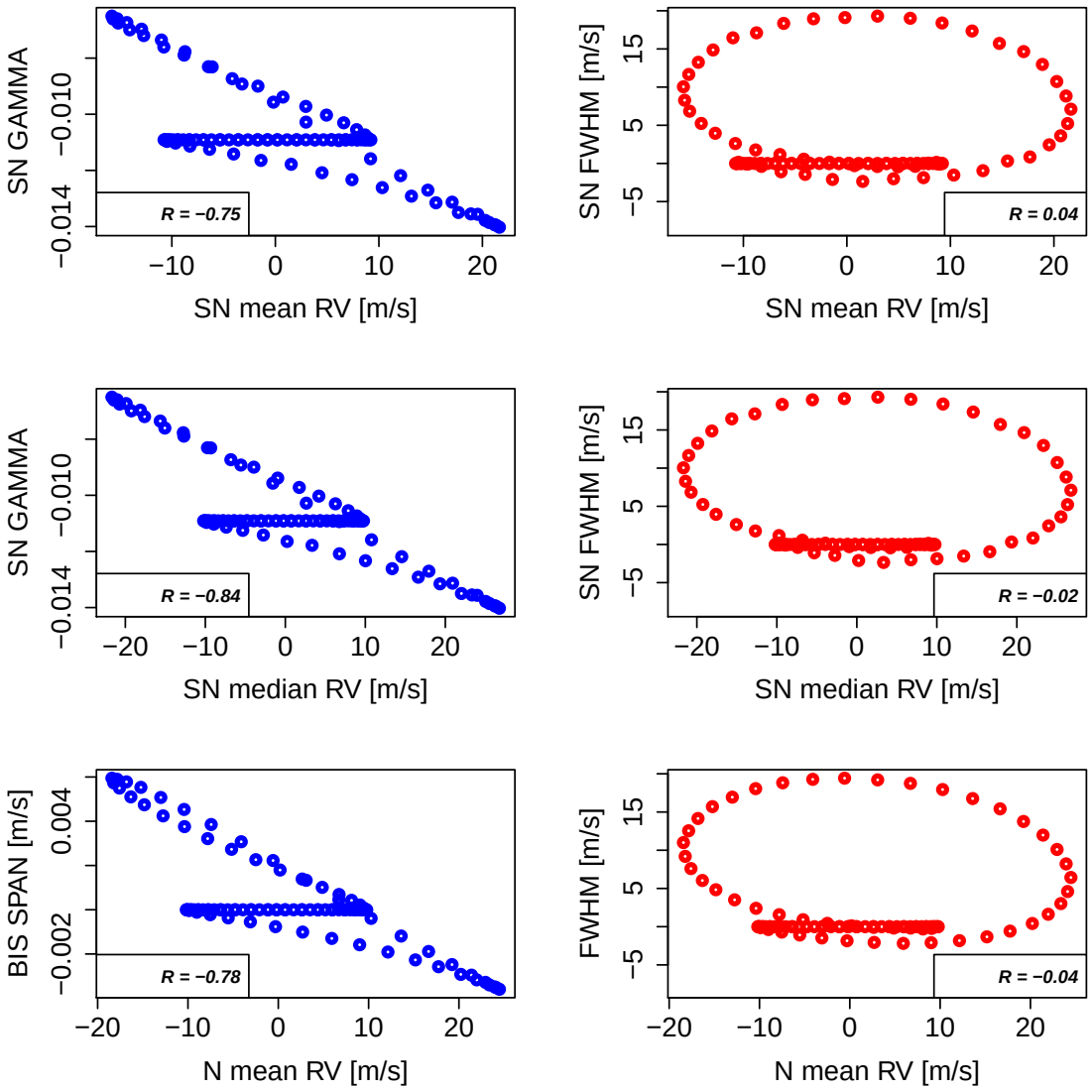

Fig. 9. Evaluation of the correlation between the RVs and the asymmetry parameters of the simulated data with a 1% spot and an injected planetary signal. The shape of the CCF changes as the spot moves, producing statistically significant correlations only between the estimated RVs and the asymmetry parameter. The correlations between the estimated RVs and the width parameter of the CCF is weaker than the case with only a spot.

of the planets with different phases were detected was found. The results are shown in Fig. 13. All three of the different RV estimators give similar detection limits. Though more investigations should be considered, these results suggest that any of the RV estimators can be used when searching for a planetary signal in RV data contaminated by stellar activity.

Although the detection limits remain relatively constant for the periods ≤15 days, there is a notable increase in the detection limit at 20 and 30 days. This is likely due to the periods of the simulated planets, 20 days and 30 days, are close to the first harmonic or to the rotational period of the star (36.7 days, Dumusque et al. 2012) and therefore close to the semi-periodicity of the stellar activity signal. In such a case, the activity correction absorbs part of the planetary signal and a stronger planetary signal is needed at these periods in order for them to be detectable.

6. Estimation of standard errors for the CCF parameters

In this section, we investigate how the noise of the CCF influences the CCF parameters derived either by a Normal density or a SN density fit. [[**Jessi:** Should we note that the CCF noise is from the photon-noise of the spectra? Is there more to "noise" in the previous sentence?]] Because a CCF is obtained from a cross-correlation, each point of a CCF is correlated with the other points. Therefore, we cannot simply vary each point in the CCF by their respective error bars and then recalculate the best SN or Normal density fit to see how the CCF noise influences the estimation of the parameters of interest (i.e., N mean RV, SN mean RV, SN median RV, FWHM, SN FWHM, BIS SPAN and γ). Instead, we go to the individual spectra where the individual measurements can be considered independent fo the other points. The standard error on each point of a spectrum is given by the photon noise, which follows a Poisson distribution and

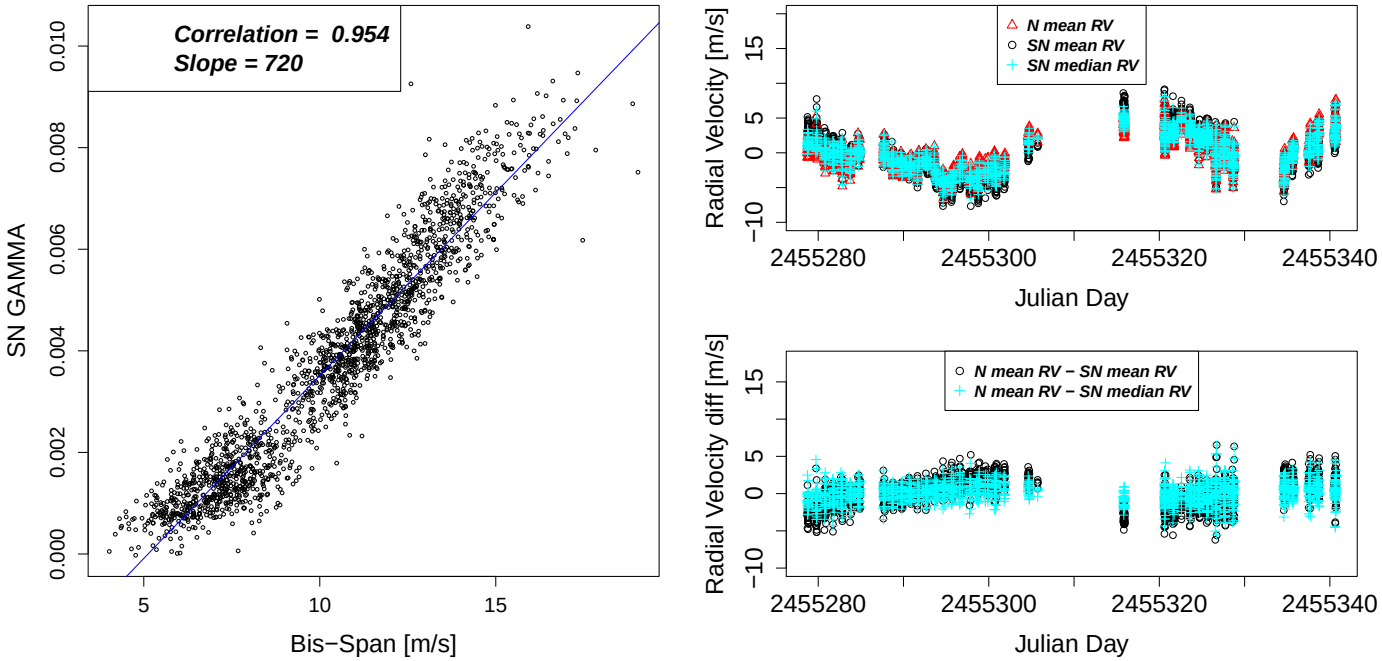

Fig. 10. (left) Correlation between γ and the BIS SPAN for Alpha Centauri B. The strong correlation suggests these two parameters are similarly measuring the asymmetry. (top right) RVs as function of Julian Day for Alpha Centauri B in 2010. The RVs are estimated using the mean of a Normal fitted to the CCF (red triangles), or the mean (black circles) or median (cyan crosses) of a SN density fitted to the CCF. (bottom right) Differences between the RVs estimated with the SN density and those from the Normal density.

is therefore estimated by taking the square-root of the measured flux.

The following method was carried out in order to estimate the error bars on the different parameters derived from the CCF. We first modify the values of all the points in the spectrum given their respective error bars. To do so random Gaussian noise centered on the value of the point and with standard deviation the square-root of the flux was added across each spectrum. The CCF was calculated using this spectrum according to the method presented in Pepe et al. (2002), then fitted either a Normal or SN density to this CCF and recorded the different parameters. This process was repeated a hundred times in order to obtain a distribution for each CCF parameter. The standard deviations of the resulting distributions provide standard errors for each CCF parameters.

The standard errors were also computed for each CCF parameter for all the CCFs of HD215152, HD192310 and Corot-7. [[**Jessi:** Why weren't the others stars included?]] [[**Umberto:** because the computational time for Tau Ceti and Alpha Cent was much longer (there are much more CCF's). can we add the the SNR 'levels' are covered with the stars we used?]] [[**Jessi:** How much longer would it take to include these? Can we run them on an HPC? At the very least we should note why they are not included]] This provides information on how the noise of each CCF parameter varies as a function of SNR, where the SNR measured at 550 nm on the original spectra varies between 10 and 500. The combined results for all the stars considered are displayed in Fig. 14. The top plots show the standard errors of the three different estimated RV's, the width, and the asymmetry estimates. Note that because BIS SPAN and γ do not have the same units, the estimated slopes of the correlation between those two parameters to transform γ in $\mathrm{m\,s^{-1}}$were used (see Fig. 10 and Table A.1 for the value of the slope for each star). In the bottom plots show the ratio between the standard errors measured when using the SN parameters and the Normal parameters. Values smaller or larger than one will imply that standard errors from the SN parameters are more or less precise than the Normal parameters, respectively.

The estimates for the RV all appear to follow a similar exponential decay. Although the results for three different stars are included in the plot, there do not appear to be any offsets in this decay, which implies that the SNR at 550 nm is the main contributor to the precision measured in RV. [[**Jessi:** The previous sentence does not make sense to me. Why would no offsets imply it is the main contributor in the precision?]] This is not surprising as the three stars studied here are all main sequence K-dwarfs.

When comparing the three different estimates for the RV, the SN mean RV presents standard errors that are 60% larger than the N mean RV. However, the SN median RV gives errors 10% more precise than N mean RV. The parameters describing the width of the CCF, FWHM and SN FWHM, have comparable standard errors. Finally, for the asymmetry parameters, γ has standard errors that are 15% more precise than BIS SPAN. In conclusion, when fitting a SN density to the CCF and using as RV estimates SN median RV, we are able to improve by 10% the precision on the estimated RV, 15% the precision on the estimated asymmetry, both without degrading the precision on the estimated width. SN mean RV should not be use to derive precise estimated RVs, except perhaps in specific conditions described below, as the precision on this parameter is 60% worse than the precision on the RVs derived from N mean RV.

7. Discussion

When fitting a SN density to the CCF, parameters that describe the RV (i.e. the CCF barycenter), the amplitude (sometimes

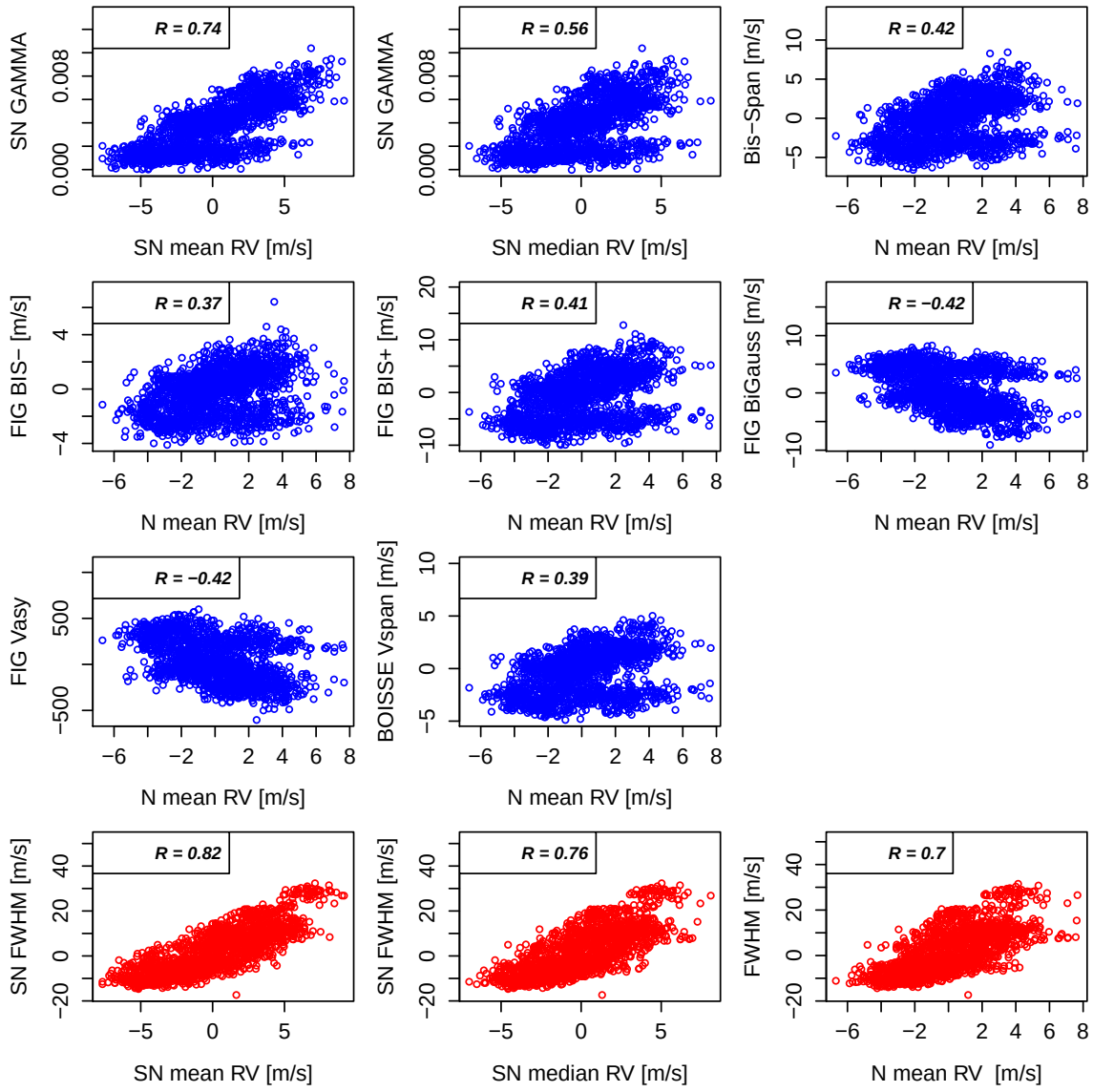

Fig. 11. (top three rows) Correlations between the asymmetry parameters and their corresponding estimated RVs for Alpha Cen B. (bottom row) The correlation between the FWHM and the estimated RVs for Alpha Centauri B. The correlations are stronger when using parameters derived from the SN fit than the Normal one. The estimated R's are all statistically significant.

called the CCF contrast), the width, and the asymmetry of the CCF can all be estimated in a single model framework. For the RV estimator, we investigated the use of the mean and the median of the SN density. Then the width is derived using the variance of the SN density (SN FWHM= $2\sqrt{2ln(2)\sigma^2}$) and the asymmetry by using γ, the skewness parameter of the SN density.

To analyze the performance the proposed SN framework, tests on both simulated and real data were carried out and compared to the usual approach of fitting a Normal density shape (for estimating the RV and FWHM) and the BIS SPAN (for estimating the asymmetry). The simulated CCFs were generated using the SOAP 2.0 code, which can simulate activity signals induced by a spot or a facula on a solar-like star. The results of the simulation study suggest that the parameters derived from the SN density fit are more sensitive to activity than the parameters obtained by the usual Normal method. In this case, sensitivity was measured using the correlation between the asymmetry and the estimated RV's, and it turned out either SN mean RV or SN median RV had a stronger correlation with gamma than the correlation between N mean RV and BIS SPAN. Moreover, the correlation between the FWHM and the estimated RV's is also stronger when using the parameters from the SN compared to the parameters from the Normal. The SN parameters continued to have stronger correlations than the Normal in the setting where a planetary signal was added the SOAP 2.0 with a single spot.

We also propose a new model to correct estimated RV data for stellar activity signals. Generally, when fitting for planetary signals, it is common to use a model composed of one or several Keplerian signals to account for the planets, in addition to a linear combination of the FWHM and BIS SPAN to account for stellar activity signals. The proposed model adds an additional term to account for the amplitude of the CCF, and an interaction term between the estimated asymmetry and the width parameters. Using the simulated data from SOAP 2.0, this new model

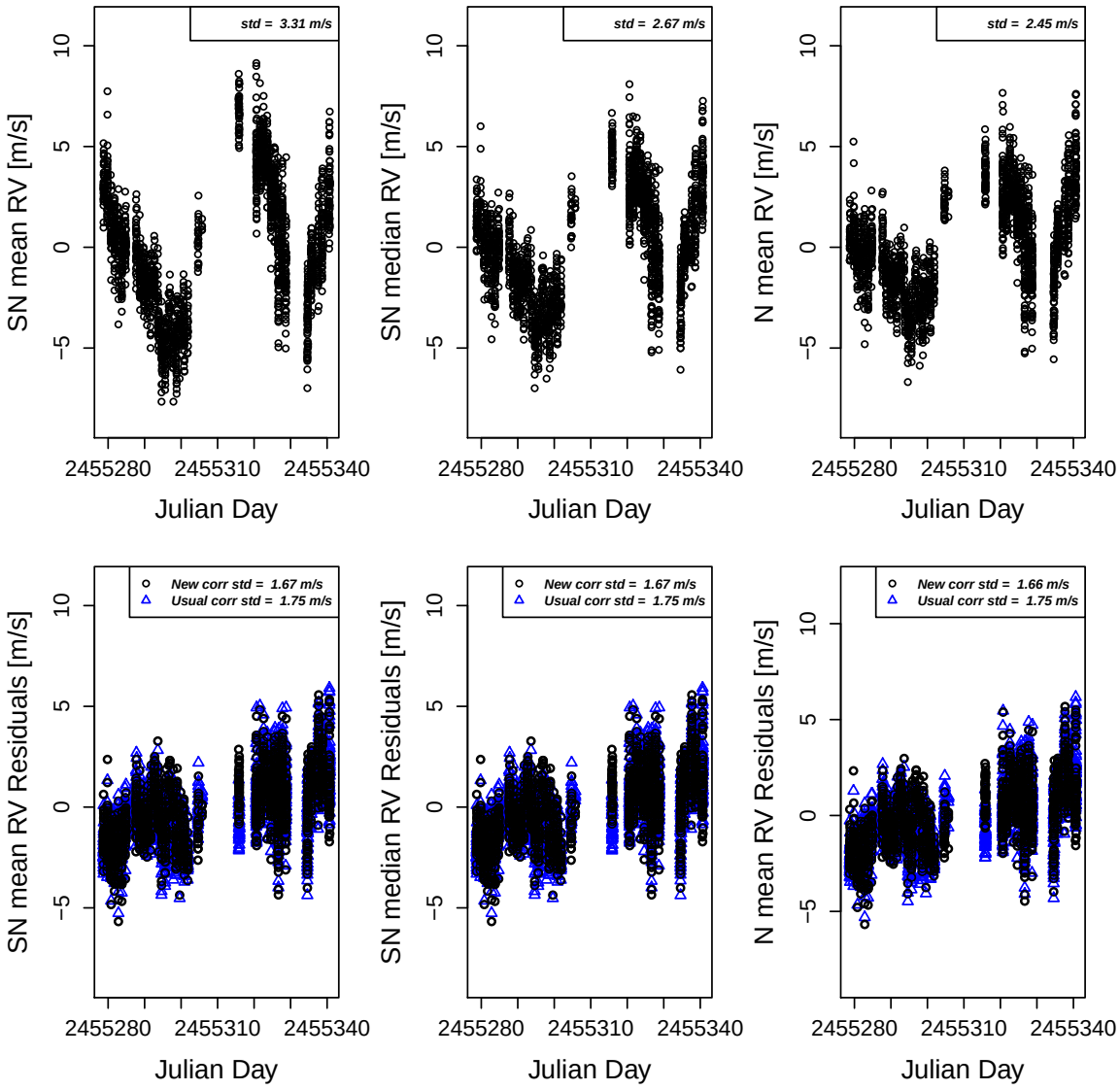

Fig. 12. (top) The RVs (black dots) for Alpha Centauri B estimated using a SN and a Normal fit. (bottom) The residuals from the model fit using Eq. 5 (New corr. std, black dots) and the residuals from the usual correction (Usual corr. std, blue triangles), based on $RV_{\text{activity}} = \beta_0 + \beta_1\gamma + \beta_2$SN FWHM for the SN fit and on $RV_{\text{activity}} = \beta_0 + \beta_1$BIS SPAN $+ \beta_2$FWHM for the normal fit. The residuals have a smaller systematic component when using the proposed model of Eq. 5 (black dots) compared to the usual model (blue triangles).

reduces the effect of the stellar activity signal by a factor of 2 and 3.5 over the usual model, for the facula and spot, respectively.

Similar conclusions were reached with the real data as with the results of the SOAP 2.0 simulations regarding the strength of the correlation between the CCF asymmetry and the RV, as well as between the CCF width and the RV. However, the simulated data stronger asymmetry-RV correlation between γ and SN median RV while the real data had stronger correlations between γ and SN mean RV. There could be several reasons for this difference. First, it may be due to the fact that in SOAP 2.0 uses a variation of a spot template spectrum as the model rather than a facula template spectrum so the SOAP spectra with a facula is not accurately modeling this type of activity. Because the temperature between a spot and a faculae is significantly different, their spectra should be different. Additionally, there are expected to be multiple active regions on a star at different locations in longitude and latitude, while the SOAP 2.0 data used in the simulation study included only a single active region. These reasons could also be responsible for the general differences observed in the behavior of the mean and the median of the SN between the SOAP 2.0 data and real data. [[**Jessi:** regarding the previous statement, would you say that there are also differences in the N mean RV generally between the SOAP and real data? I wonder if that could somehow be tied in?]] Nevertheless, in both the simulated and the real cases, the parameters derived from the SN are generally more sensitive to activity than the parameters derived from the Normal. Also, the correlation between the asymmetry and the SN mean RV is always stronger than the parametrization of the CCF presented in Boisse et al. (2011) and Figueira et al. (2013). Since if an apparent RV signal is induced by activity generally a strong correlation will be observed between the RV and the FWHM of the CCF or its BIS SPAN, to use the SN density to fit the CCF leads to a better understanding of the spurious variations in RV caused by stellar activity.

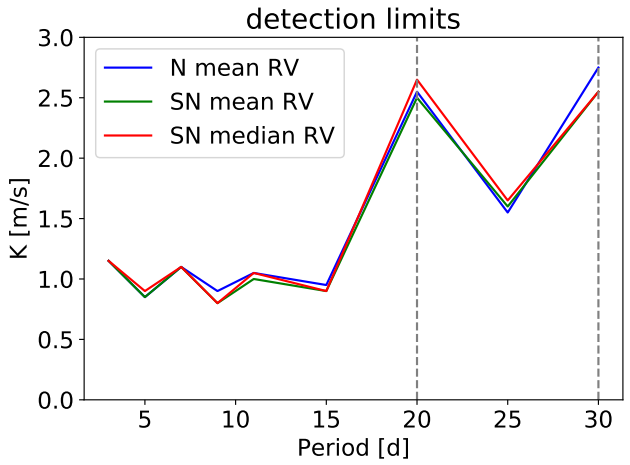

Fig. 13. Detection limits of planetary signals once the stellar activity signal is removed from the raw RVs using the model proposed in Eq. 5. [[**Jessi:** For the black-and-white versions, it would be helpful to have different line types plotted]] The different RV estimators have similar detection limits.

Considering the different RV estimates of the real data, the amplitude of stellar activity tends to be largest for the SN mean RV, followed by the SN median RV, and then N mean RV. Although there appears to be a larger difference between the SN mean RV and the SN median RV, the SN median RV is similar to the N mean RV. Hence, the mean of the SN density appears to be more sensitive to variation in the CCF shape than the median of the SN or the mean of the Normal.

Having an estimator of RV that is more sensitive to stellar activity, such as the SN mean RV, can help with better probing stellar rotational periods or better understanding the covariance of stellar signals when fitting a Gaussian Process to the RVs (e.g. Faria et al. 2016; Haywood et al. 2014). We saw in the preceding section that the SN mean RV estimator is 60% noisier than the N mean RV estimator. This is not necessarily a negative aspect. Indeed, in the case were photon-noise is not dominant, stellar activity can be better characterized if its effect can be amplified by measuring the SN mean RV, even if the uncertainties related to this parameter are increased by 60%.

Even if the different RV estimators generally have different amplitudes, once the new correction for stellar activity is applied, the residuals of the model have similar rms. When comparing the activity correction proposed in this paper with the usual correction that only uses a linear combination of the CCF asymmetry and width, for the simulations based on the presence of a facula or a spot the new proposed correction almost entirely explains the spurious variations in RV. However, when moving to real data, there is just a slight improvement by using the proposed correction function for stellar activity. Anyway, we encourage the community to use this new simple linear model if the goal is to reduce activity without considering the computationally expensive use of Gaussian Process. A further analysis of the p–value for each coefficient can be performed to see the ones that are not relevant and that can therefore be removed.

A test was carried out to see if some RV estimates were better at finding planets in RV data affected by observed stellar signals. When using the new correction proposed in this paper to mitigate the effect from stellar activity and when assessing planetary detection based on a periodogram analysis, we find very similar detection limits for the three RV estimators presented in this paper. [[**Jessi:** Maybe to try another test to see if the proposed correction is helpful, the detection-limit test could be carried out with the usual correction and we can see if it does not performa s well.]] Therefore it seems that any of the RV estimate can be used to search for planetary signals.

Finally, we investigated the precision of each of the SN and Normal parameters including the BIS SPAN. It turns out that the SN mean RV should not be used to get precise RVs as the standard errors on this parameter is 60% greater than the N mean RV. However, the SN median RV is 10% more precise than the N mean RV. Regarding the asymmetry estimates, we observe that γ has a precision 15% better than the BIS SPAN.

8. Conclusion

When searching for low-mass exoplanets using the RV technique, it is necessary to get precise estimates of the RV and also to account for variations induced by stellar activity in order to avoid false detections. Stellar activity such as spots and faculae can lead to shape variations in the spectra features, which then results in shape variations of the CCF. The correlations between the width or asymmetry of the CCF and the estimated RV are commonly used as way to detect if the RVs are affected by stellar activity signals. Because the presence of real planets would result in only a shift in the CCF (not a shape change), strong correlations between the shape features of the CCF and the estimated RVs suggest that stellar activity may be present.

In this paper, a new approach for measuring shape changes in the CCF is proposed using the SN density to estimate the RV and shape of the CCF. This new method is compared to a commonly used method based on a Normal density fit to the CCF. The mean of the Normal density is used as the estimated RV and the FWHM estimates the width of the CCF. Because the Normal density does not have any skewness, another method is need to estimate the asymmetry of the CCF, and the BIS SPAN is often employed. In addition, the proposed SN approach is compared to other parameterizations of the CCF asymmetry that have been shown to be sensitive to activity signals (Boisse et al. 2011; Figueira et al. 2013).

In the different tests carried out for this work, the SN parameters tended to perform at least as well (and sometimes better) as the parameters from the Normal approach with the BIS SPAN. The SN parameters γ, SN FWHM and SN mean RV always show stronger correlation than the correlations between any of the parameters derived by the Normal and the BIS SPAN, or the different asymmetry parametrization presented in Boisse et al. (2011) and Figueira et al. (2013). This suggests the SN parameters may be better at probing stellar activity signals than the other methods. In addition, the precision measured on the SN median RV and the γ are 10 and 15% more precise than the N mean RV and the BIS SPAN, respectively.

Because of the advantages of using the proposed SN approach compared to the Normal approach, BIS SPAN, and all the asymmetry parameters described in Boisse et al. (2011) and Figueira et al. (2013), this new parametrization of the CCF may be more useful for detecting stellar activity than the previously proposed parametrizations. Correlations between γ and the SN mean RV, and between the width and the SN mean RV can be used to probe stellar activity signals in RV data, and the SN median RV can be used to estimate RV.

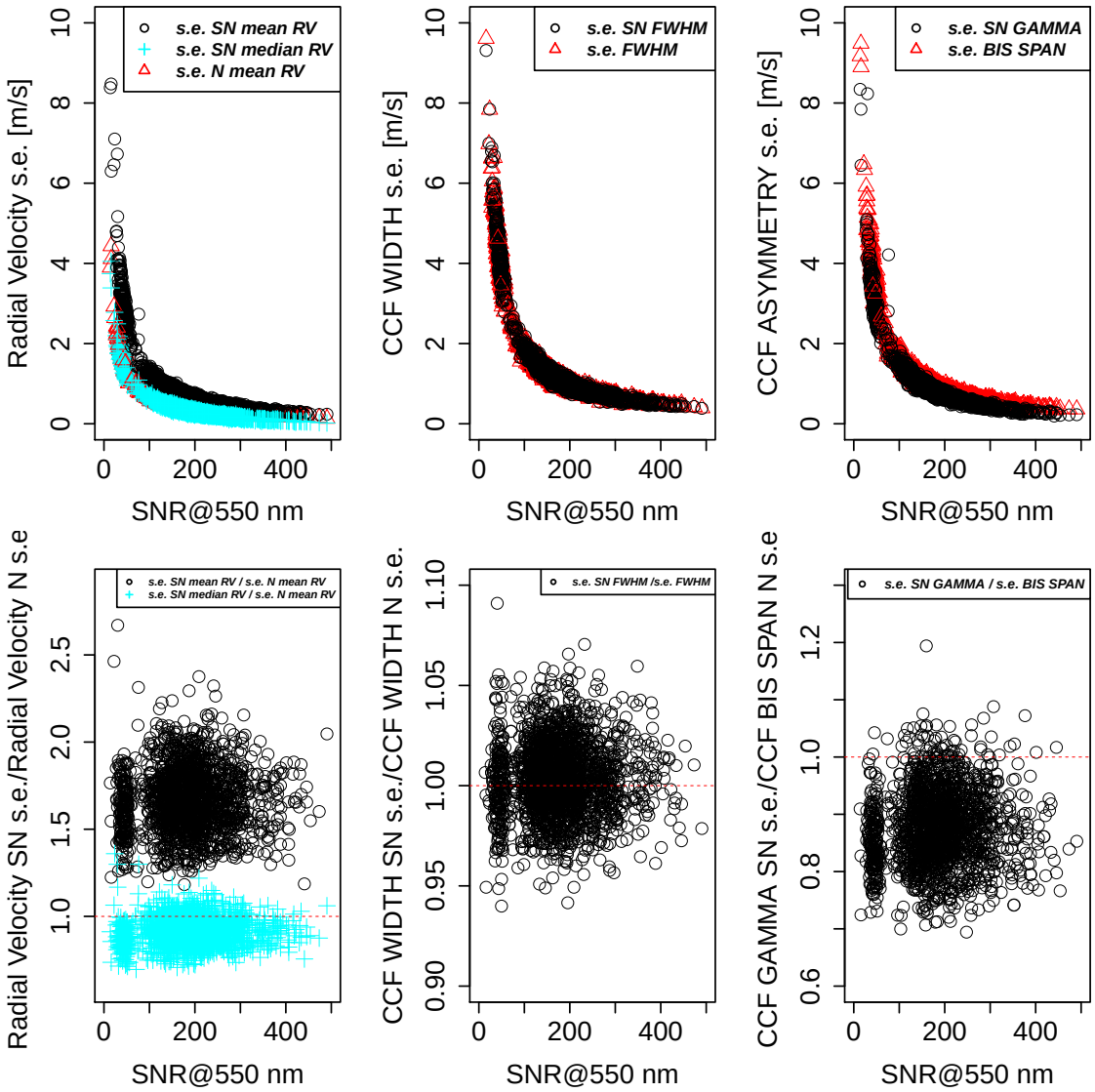

Fig. 14. *Bootstrap analyses on the stars HD215152, HD192310 and Corot-7.* (top) Comparison between the standard errors from the bootstrap analysis of the estimated RVs, FWHM, and asymmetry parameters using the SN fit and the common strategy (Normal fit and BIS SPAN). (bottom) Ratio between the standard errors retrieved on the parameters derived from the common strategy and the corresponding standard errors retrieved on the parameters derived from the SN fit. When using SN mean RV (black circles), the standard errors are in average 60% larger than the standard errors of N mean RV (red triangles). However, the standard errors for SN median RV (cyan crosses) are on average 10% smaller than the standard errors coming from the N mean RV. The use of the asymmetry SN parameter γ leads to standard errors in average 15% smaller than the standard errors related to the BIS SPAN. Note that for the asymmetry, the error in BIS SPAN is in $\mathrm{m\,s^{-1}}$. To be able to compare the errors in γ and BIS SPAN, we multiplied the error in γ by the slope of the correlation between γ and BIS SPAN.

9. Acknowledgements

We are grateful to all technical and scientific collaborators of the HARPS Consortium, ESO Headquarters and ESO La Silla who have contributed with their extraordinary passion and valuable work to the success of the HARPS project. XD is grateful to The Branco Weiss Fellowship–Society in Science for its financial support.

Appendix A: Appendix

In this Appendix, a similar analysis as presented in Sec. 5 is discussed for four main-sequence stars: HD192310 (K2V), HD10700 (G8V), HD215152 (K3V) and finally Corot-7 (K0V). [[**Jessi:** Are there papers that can/should be cited here for the datasets used?]]

Table A.1 summarizes the results obtained by the SN and Normal models. These results are consistent with those from the analysis of Alpha Centauri B. The correlation between γ and SN mean RV is stronger than the correlation between the BIS SPAN and N mean RV or between the asymmetry parameters derived in Boisse et al. (2009) and Figueira et al. (2013) and N mean RV for all the considered stars. The correlation between SN FWHM and SN mean RV is stronger than the correlation between FWHM and N mean RV for three of the four stars. Also for all these stars the originally estimated RVs were corrected from spurious variations in RVs caused by stellar activity using Eq. 5. Fig. A.2–A.6 show the resulting corrected RVs. The stel-

lar activity-corrected Normal and SN residuals are comparable for the stars HD192310 and HD10700. However, the rms of the residuals for HD215152 are 0.05 $m\,s^{-1}$lower for the SN model than the Normal model, and the rms of the residuals for Corot-7 are 0.34 $m\,s^{-1}$lower. HD215152 and Corot-7 have lower SNR than the other two analyzed stars. [[**Jessi:** was the SNR of the four stars stated anywhere? If not, if we want to keep the last point, we should note what the SNR of the stars are.]]

References

Anglada-Escudé, G. & Butler, R. P. 2012, The Astrophysical Journal Supplement Series, 200, 15
Arellano, R. B. & Azzalini, A. 2010
Azzalini, A. 1985, Scandinavian journal of statistics, 171
Azzalini, A. & Capitanio, A. 2014, The skew-normal and related families. Institute of Mathematical Statistics Monographs
Baranne, A., Queloz, D., Mayor, M., et al. 1996, Astronomy and Astrophysics Supplement Series, 119, 373
Belsley, D. A. 1991, Conditioning diagnostics: Collinearity and weak data in regression No. 519.536 B452 (Wiley New York)
Boisse, I., Bouchy, F., Hébrard, G., et al. 2011, Astronomy & Astrophysics, 528, A4
Boisse, I., Moutou, C., Vidal-Madjar, A., et al. 2009, Astronomy & Astrophysics, 495, 959
Borgniet, S., Meunier, N., & Lagrange, A.-M. 2015, A&A, 581, A133
Claret, A. & Bloemen, S. 2011, A&A, 529, A75
Cosentino, R., Lovis, C., Pepe, F., et al. 2012, in Proc. SPIE, Vol. 8446, 84461V
Desort, M., Lagrange, A.-M., Galland, F., Udry, S., & Mayor, M. 2007, Astronomy & Astrophysics, 473, 983
Dumusque, X. 2014, ApJ, 796, 133
Dumusque, X. 2016, Astronomy & Astrophysics, 593, A5
Dumusque, X., Boisse, I., & Santos, N. 2014, The Astrophysical Journal, 796, 132
Dumusque, X., Borsa, F., Damasso, M., et al. 2017, Astronomy & Astrophysics, 598, A133
Dumusque, X., Pepe, F., Lovis, C., et al. 2012, Nature, 491, 207
Dumusque, X., Udry, S., Lovis, C., Santos, N. C., & Monteiro, M. 2011, Astronomy & Astrophysics, 525, A140
Faria, J., Haywood, R., Brewer, B., et al. 2016, Astronomy & Astrophysics, 588, A31
Feng, F., Tuomi, M., & Jones, H. R. 2017, Astronomy & Astrophysics, 605, A103
Figueira, P., Santos, N., Pepe, F., Lovis, C., & Nardetto, N. 2013, Astronomy & Astrophysics, 557, A93
Fischer, D. A., Anglada-Escude, G., Arriagada, P., et al. 2016, Publications of the Astronomical Society of the Pacific, 128, 066001
Hatzes, A. P. 2002, Astronomische Nachrichten, 323, 392
Haywood, R., Collier Cameron, A., Queloz, D., et al. 2014, Monthly notices of the royal astronomical society, 443, 2517
Kurster, M., Endl, M., Rouesnel, F., et al. 2003, ASTRONOMY AND ASTROPHYSICS-BERLIN-, 403, 1077
Lagrange, A.-M., Desort, M., & Meunier, N. 2010, Astronomy & Astrophysics, 512, A38
Lindegren, L. & Dravins, D. 2003, Astronomy & Astrophysics, 401, 1185
Lomb, N. R. 1976, Astrophysics and Space Science, 39, 447
Mayor, M., Pepe, F., Queloz, D., et al. 2003, The Messenger, 114, 20
Meunier, N., Desort, M., & Lagrange, A.-M. 2010, Astronomy & Astrophysics, 512, A39
Oshagh, M., Boisse, I., Boué, G., et al. 2013, A&A, 549, A35
Pepe, F., Mayor, M., Galland, F., et al. 2002, Astronomy & Astrophysics, 388, 632
Pepe, F., Molaro, P., Cristiani, S., et al. 2014, Astronomische Nachrichten, 335, 8
Queloz, D., Bouchy, F., Moutou, C., et al. 2009, Astronomy & Astrophysics, 506, 303
Queloz, D., Henry, G., Sivan, J., et al. 2001, Astronomy & Astrophysics, 379, 279
Rajpaul, V., Aigrain, S., Osborne, M. A., Reece, S., & Roberts, S. 2015, Monthly Notices of the Royal Astronomical Society, 452, 2269
Robertson, P., Mahadevan, S., Endl, M., & Roy, A. 2014, Science, 1253253
Saar, S. H. & Donahue, R. A. 1997, The Astrophysical Journal, 485, 319
Scargle, J. D. 1982, ApJ, 263, 835
Thompson, A., Watson, C., de Mooij, E., & Jess, D. 2017, Monthly Notices of the Royal Astronomical Society: Letters, 468, L16
Zechmeister, M. & Kürster, M. 2009, A&A, 496, 577

Table A.1. Notable correlations between the asymmetry or the FWHM parameters and the RVs for four stars: HD192310, HD10700, HD215152 and Corot 7. The complete results of the analyses of the correlations for the four stars are presented in Fig. A.1–A.7.

Star	# CCFs	R(SN γ, Bis-Span)	slope(SN γ, Bis-Span)	R(SN γ, SN mean RV)
HD192310	1577	0.888	786	0.669(0.64; 0.695)
HD10700	7928	0.78	604	0.322(0.302; 0.342)
HD215152	273	0.763	794	0.571(0.485; 0.646)
Corot 7	173	0.814	607	0.561(0.450; 0.656)

Star	R(Bis-Span, N mean RV)	R(FIG BiGaussian, N mean RV)	R(SN FWHM, SN mean RV)	R(FWHM, N mean RV)
HD192310	0.329(0.285; 0.373)	−0.333(−0.376; −0.289)	0.666(0.637; 0.692)	0.476(0.4367; 0.514)
HD10700	−0.073(−0.095; −0.0051)	0.127(0.105; 0.148)	0.421(0.403; 0.439)	0.529(0.513; 0.545)
HD215152	−0.067(−0.184; 0.052)	0.269(0.155; 0.376)	0.210(0.094; 0.321)	−0.138(−0.253; −0.020)
Corot 7	0.092(−0.058; 0.238)	−0.335(−0.228; −0.082)	−0.709(0.626; 0.776)	0.595(0.489; 0.683)

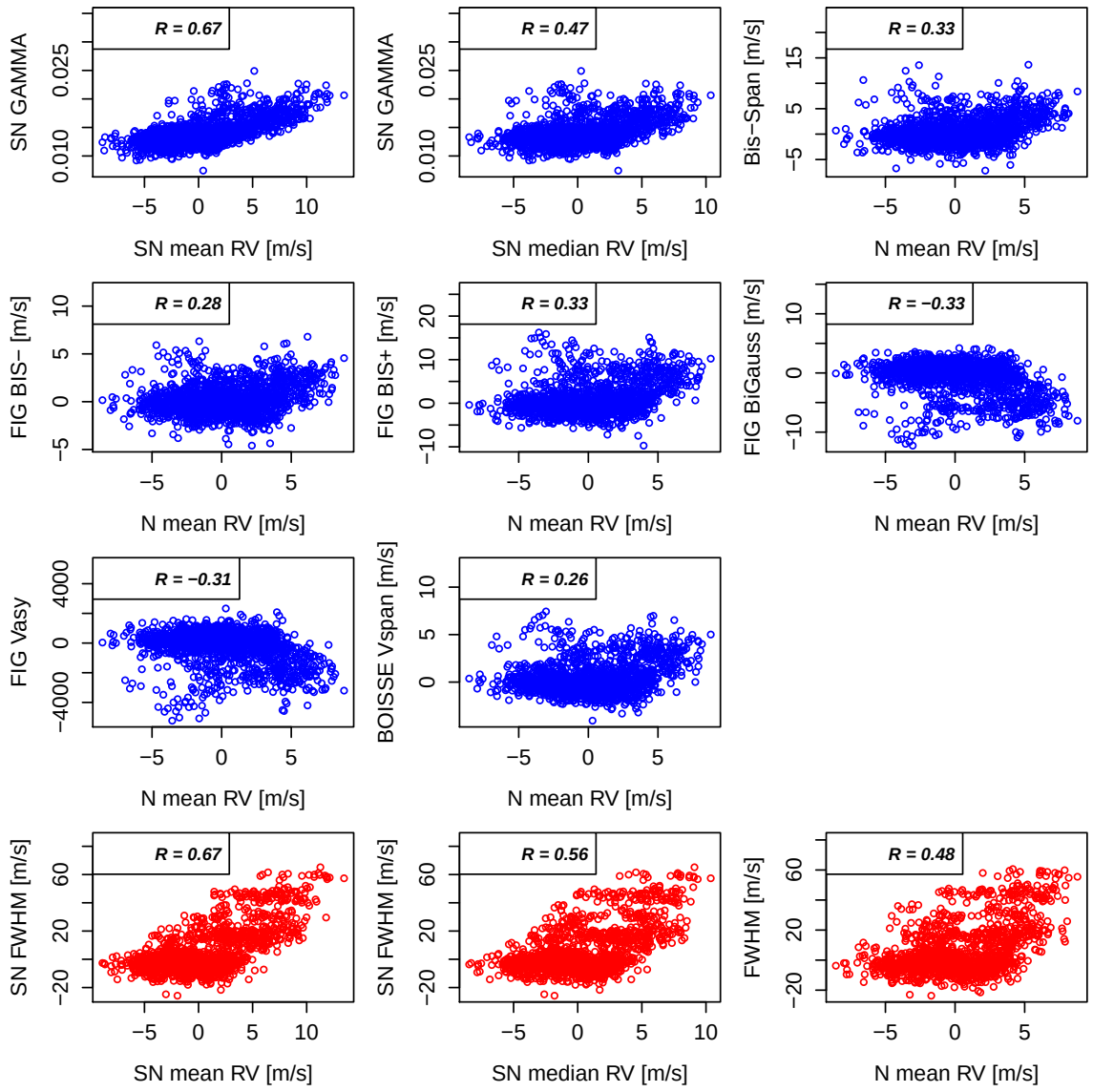

Fig. A.1. (top three rows) Correlations between the asymmetry parameters and their corresponding RVs for HD192310. (bottom row) Correlations between the FWHM and the estimated RVs. The correlations are consistently stronger when using parameters derived from the SN than the Normal. The estimated R are all statistically significant.

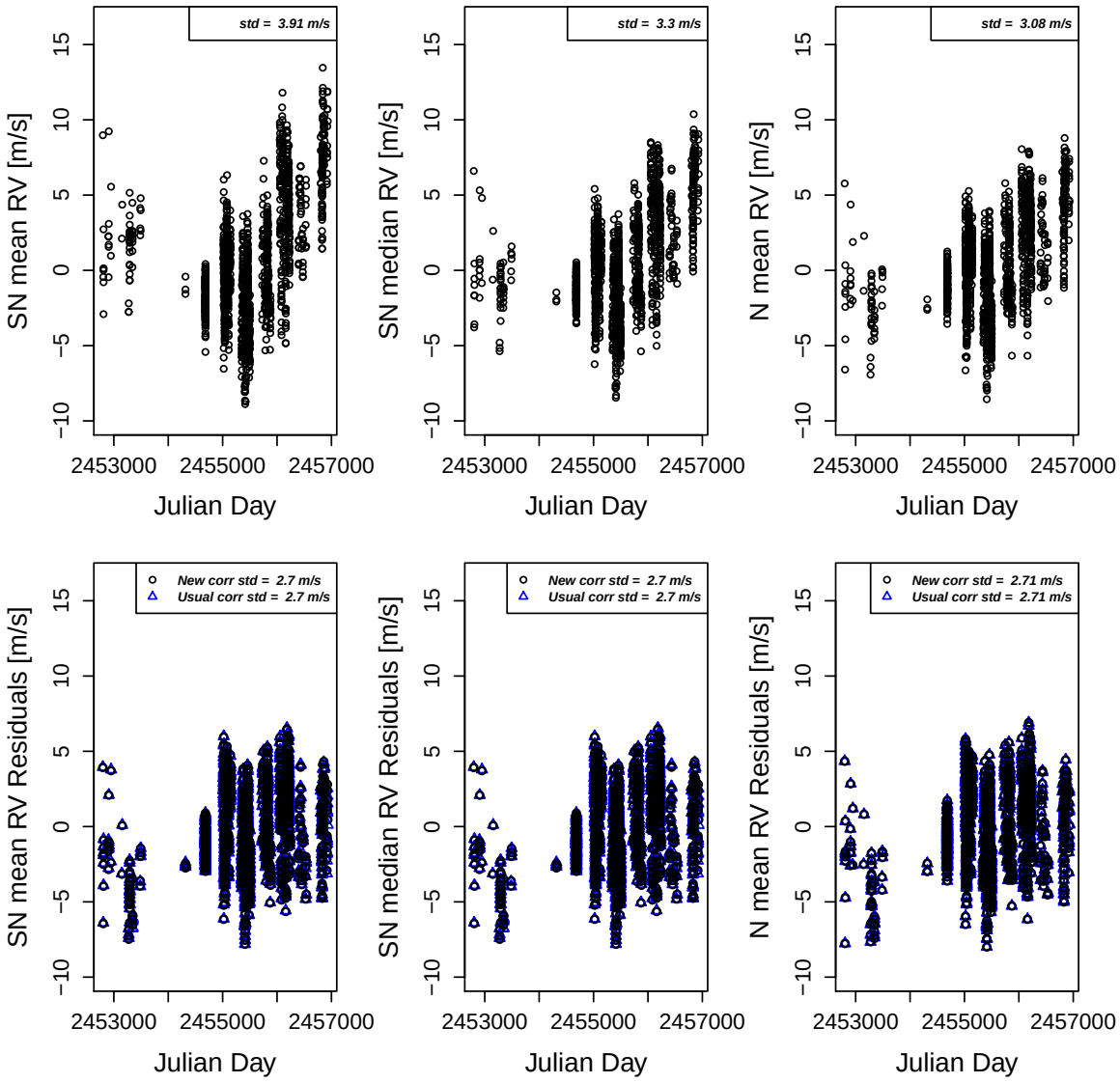

Fig. A.2. (top) The RVs (black dots) for HD192310 estimated using a SN and a Normal fit. (bottom) The residuals from the model fit using Eq. 5 (New corr. std–black dots) and the residuals from the usual correction (Usual corr. std–blue triangles), based on $RV_{\text{activity}} = \beta_0 + \beta_1\gamma + \beta_2$SN FWHM for the SN fit and on $RV_{\text{activity}} = \beta_0 + \beta_1$BIS SPAN $+ \beta_2$FWHM for the normal fit. The residuals for both the proposed correction from stellar activity are comparable.

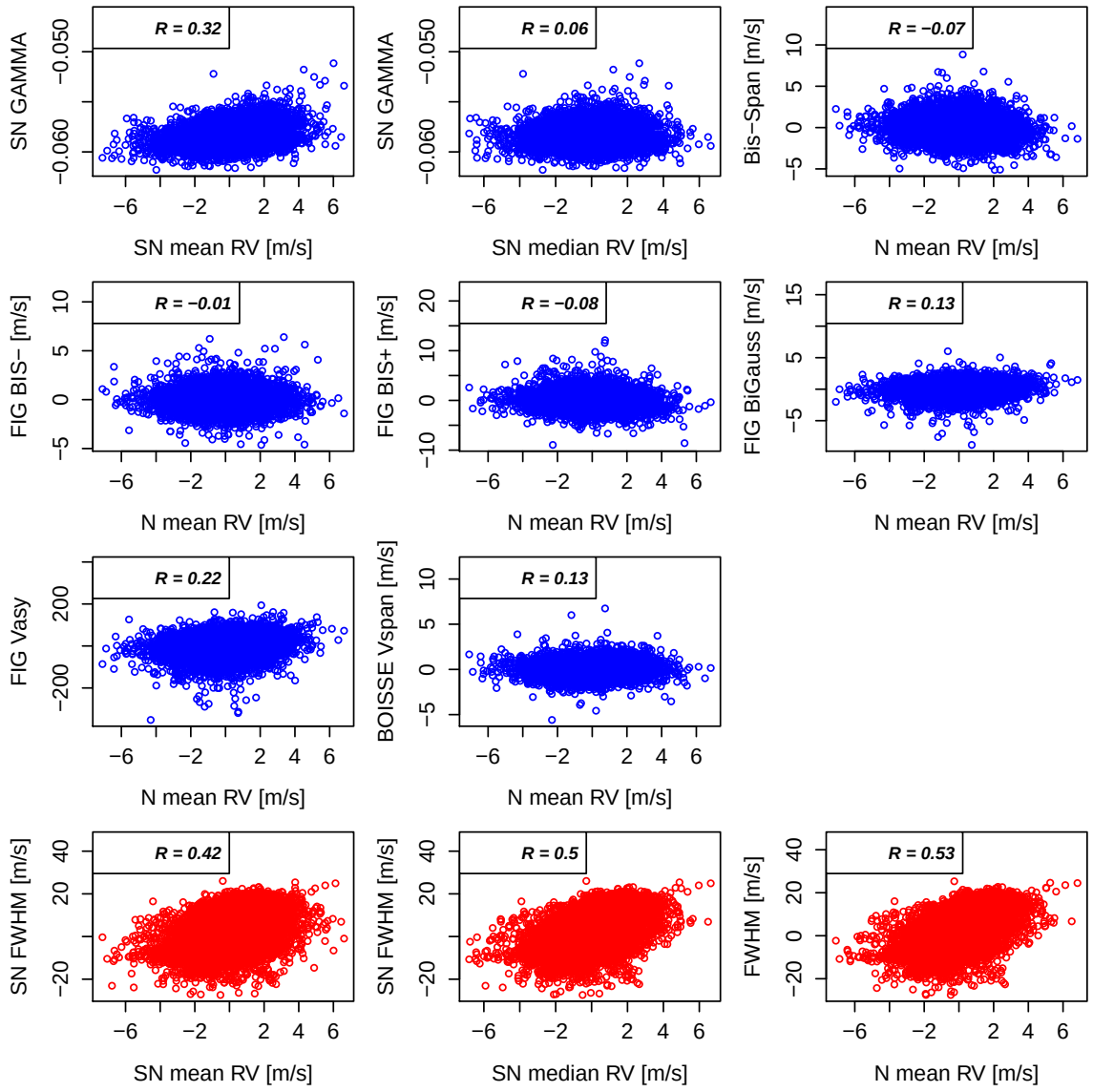

Fig. A.3. (top three rows) Correlations between the asymmetry parameters and their corresponding RVs for HD10700. (bottom row) Correlations between the FWHM and the RVs for HD10700. The correlations are consistently stronger when using SN mean RV compared to N mean RV for the asymmetry parameters; however, the correlation between the FWHM and the N mean RV, only for this quiet star, is stronger the the analogous correlations with the estimated SN RVs. The estimated R are statistically significant, except for the correlation between FIG BIS and RV (p–values=0.36).

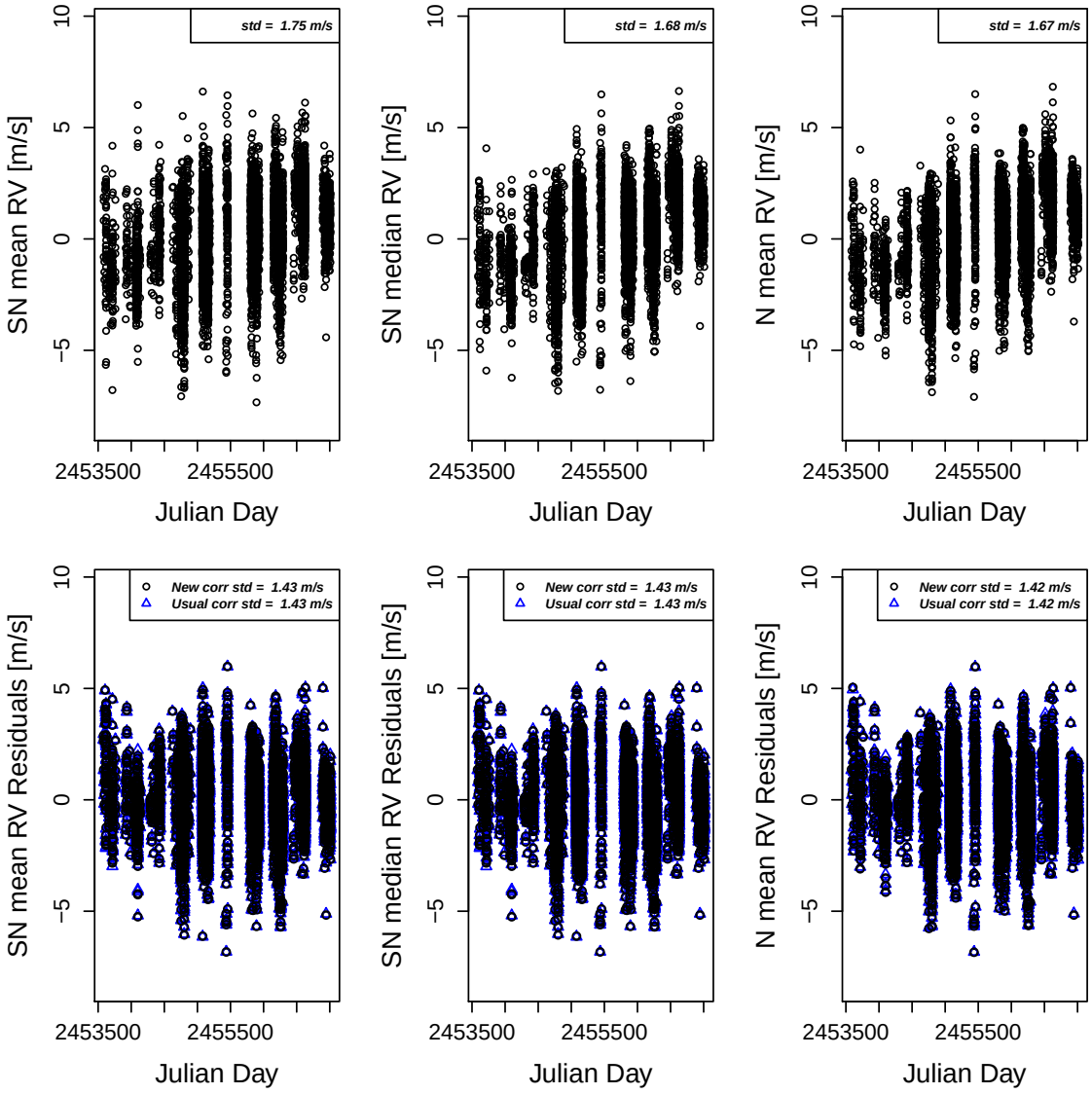

Fig. A.4. (top) The RVs (black dots) for HD10700 estimated using a SN and a Normal fit. (bottom) The residuals from the model fit using Eq. 5 (New corr. std–black dots) and the residuals from the usual correction (Usual corr. std–blue triangles), based on $RV_{\text{activity}} = \beta_0 + \beta_1\gamma + \beta_2$SN FWHM for the SN fit and on $RV_{\text{activity}} = \beta_0 + \beta_1$BIS SPAN $+ \beta_2$FWHM for the normal fit. The residuals for both the proposed correction from stellar activity are comparable.

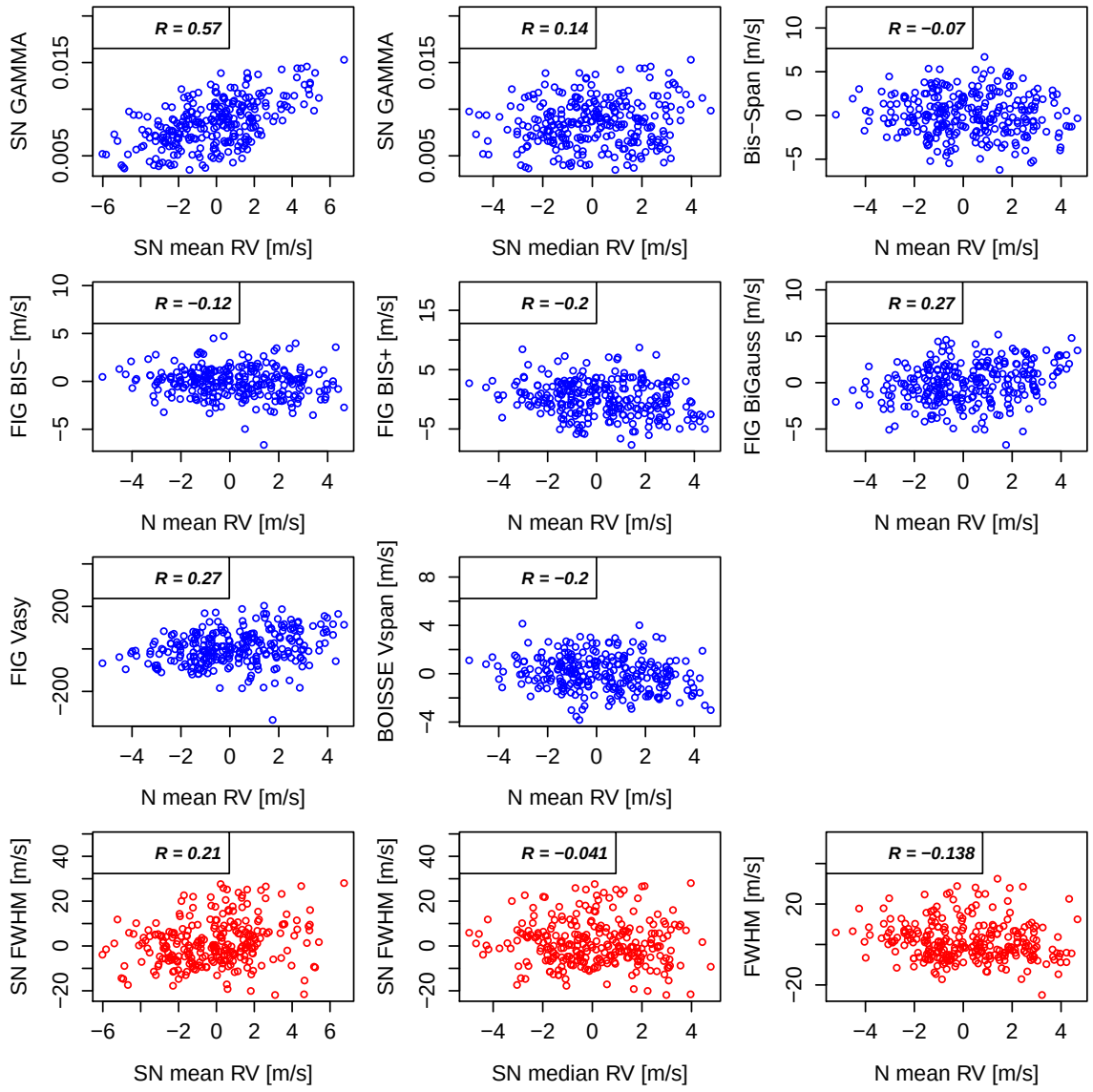

Fig. A.5. (top three rows) Correlations between the asymmetry parameters and their corresponding RVs for HD215152. (bottom row) Correlations between the FWHM and the RVs for HD215152. The correlations are consistently stronger when using SN mean RV compared to N mean RV. The p–values associated with each *R* are not statistically significant for the correlation between N mean RV and BIS SPAN (p–values=0.27), the correlation between N mean RV and FIG BIS- (p–values=0.05), the correlation between SN median RV and SN FWHM (p–values=0.5) and the correlation between N mean RV and FWHM (p–values=0.2).

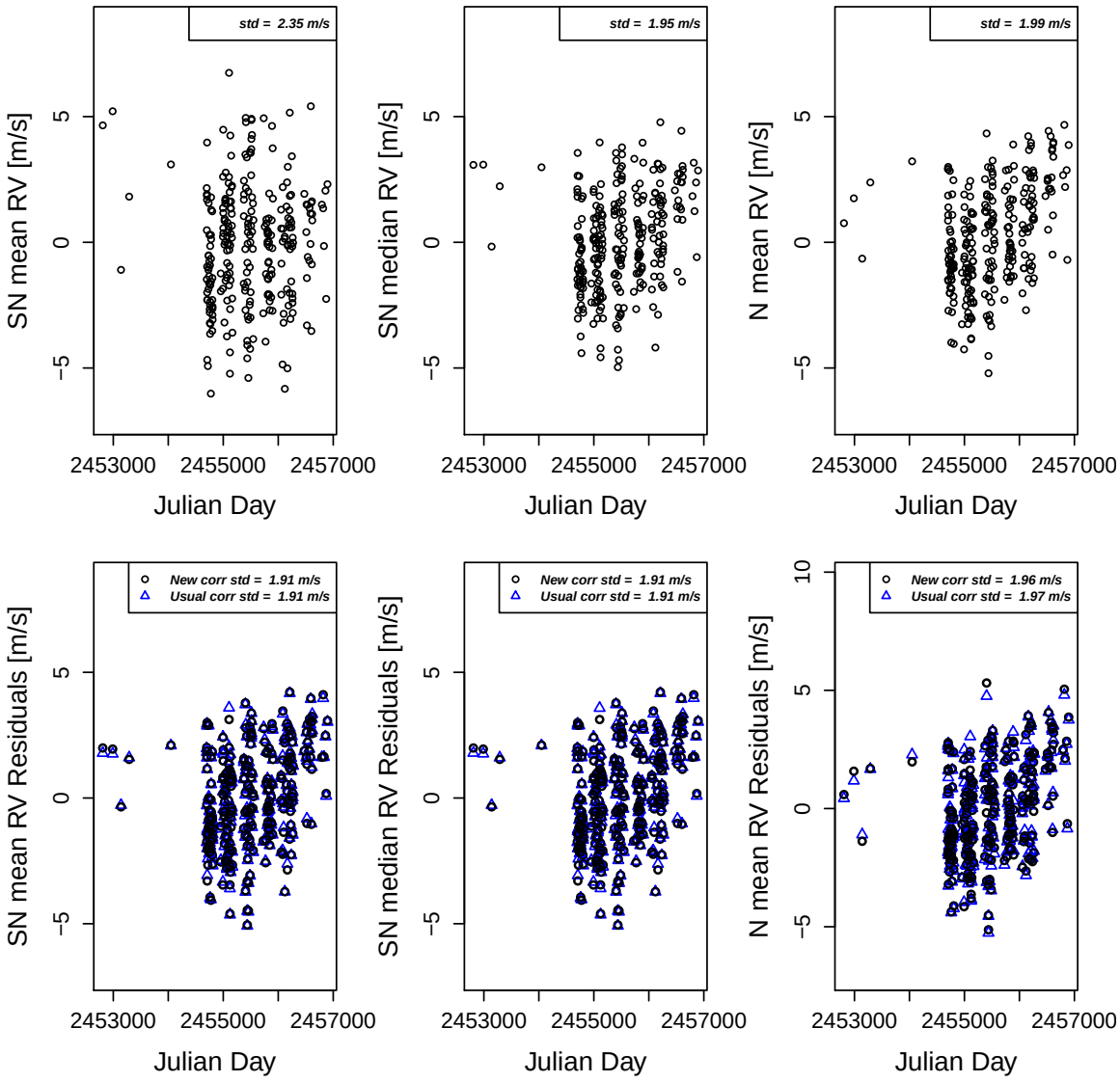

Fig. A.6. (top) The RVs (black dots) for HD215152 estimated using a SN and a Normal fit. (bottom) The residuals from the model fit using Eq. 5 (New corr. std–black dots) and the residuals from the usual correction (Usual corr. std–blue triangles), based on $RV_{\text{activity}} = \beta_0 + \beta_1\gamma + \beta_2$SN FWHM for the SN fit and on $RV_{\text{activity}} = \beta_0 + \beta_1$BIS SPAN $+ \beta_2$FWHM for the normal fit. The residuals for both the proposed correction from stellar activity are comparable.

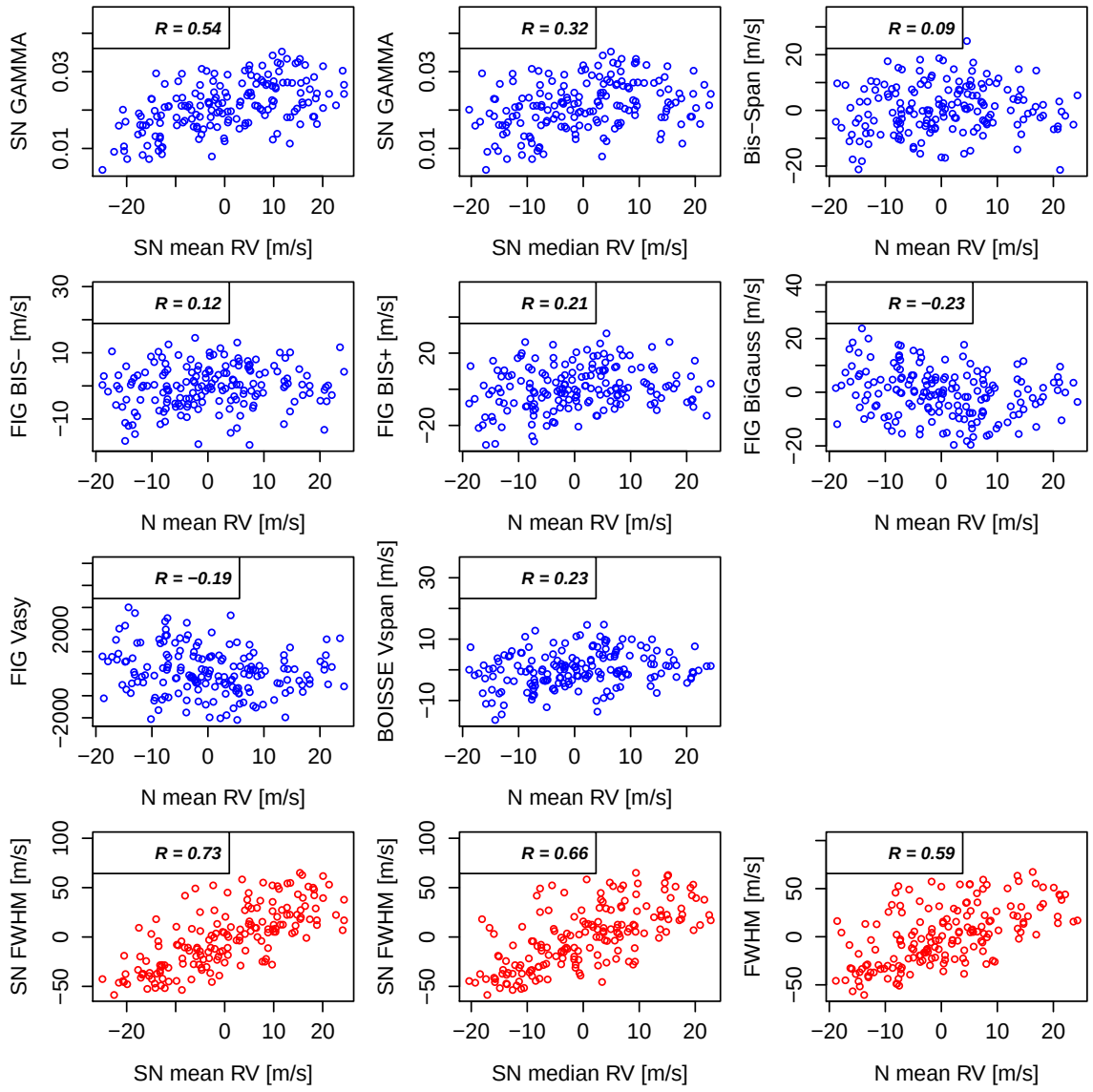

Fig. A.7. (top three rows) Correlations between the asymmetry parameters and their corresponding RVs for Corot 7. (bottom row) Correlations between the FWHM and the RVs for Corot 7. The correlations are consistently stronger when using parameters derived from the SN than the Normal. The p–values associated with each *R* are not statistically significant for the correlation between N mean RV and BIS SPAN (p–values=0.23) and the correlation between N mean RV and FIG BIS- (p–values=0.11).

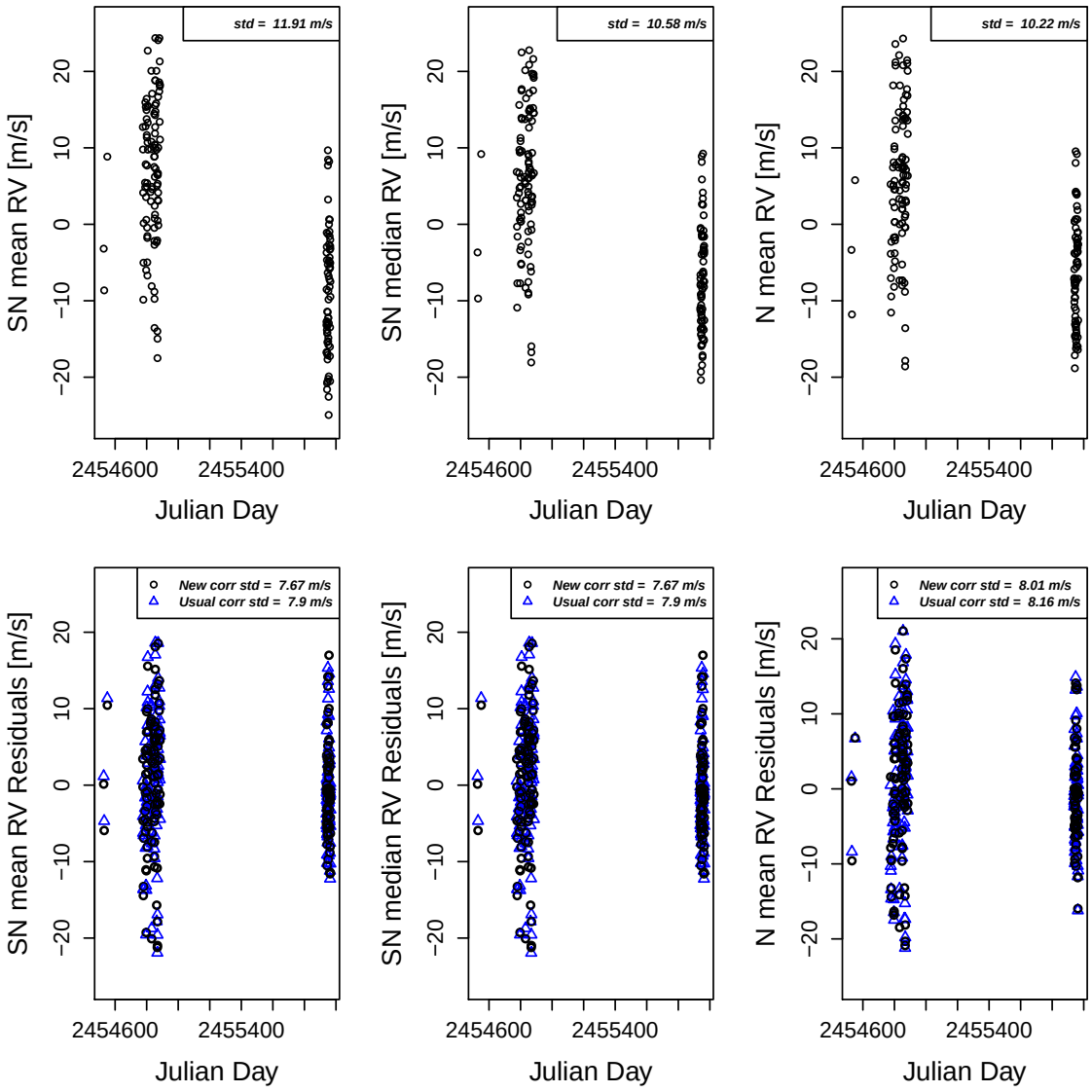

Fig. A.8. (top) The RVs (black dots) for Corot 7 estimated using a SN and a Normal fit. (bottom) The residuals from the model fit using Eq. 5 (New corr. std–black dots) and the residuals from the usual correction (Usual corr. std–blue triangles), based on $RV_{\text{activity}} = \beta_0 + \beta_1\gamma + \beta_2$SN FWHM for the SN fit and on $RV_{\text{activity}} = \beta_0 + \beta_1$BIS SPAN $+ \beta_2$FWHM for the normal fit. The residuals have a smaller systematic component when using the proposed model of Eq. 5 (black dots) compared to the usual model (blue triangles). Moreover, once corrected for stellar activity using Eq. 5, the remaining std from the SN models are 0.334 m s^{-1}smaller than the remaining std of the Normal model.